```
OM of: US-09-810-936-304 to: GenEmbl:*   out_format : pfs

Date: Aug 26, 2002  4:26 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlh
 -Q=/cgn2_1/USPTO_spool/US09810936/runat_23082002_145849_26171/app_query.fasta_1.450
 -DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
 -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
 -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
 -USER=US09810936_@CGN1_1_4255 -NCPU=6 -ICPU=3 -LONGLOG
 -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-810-936-304
Query length: 384
Database: GenEmbl:*
Database sequences: 1797656
Database length: 1873333701
Search time (sec): 1851.490000


score_list:
Sequence            Strd Orig    ZScore    EScore Len  ! Documentation  ..
gb_pat:AX106592       + 2064.00 2864.74  3.3e-151  1155  ! AX106592 Sequence 373 from Pat
gb_pat:AX140883       + 2064.00 2864.74  3.3e-151  1155  ! AX140883 Sequence 373 from Pat
gb_pat:AX200743       + 2064.00 2864.74  3.3e-151  1155  ! AX200743 Sequence 373 from Pat
gb_pat:AX267399       + 2064.00 2864.74  3.3e-151  1155  ! AX267399 Sequence 373 from Pat
gb_pat:AX282956       + 2064.00 2864.74  3.3e-151  1155  ! AX282956 Sequence 5 from Paten
gb_pat:AX316964       + 2064.00 2864.74  3.3e-151  1155  ! AX316964 Sequence 301 from Pat
gb_pat:AX316986       + 2064.00 2861.80  4.8e-151  1590  ! AX316986 Sequence 323 from Pat
gb_pat:AX316991       + 2054.00 2850.82  2.0e-150  1155  ! AX316991 Sequence 328 from Pat
gb_pat:AX106593       + 2024.00 2804.04  8.0e-148  2000  ! AX106593 Sequence 374 from Pat
gb_pat:AX140884       + 2024.00 2804.04  8.0e-148  2000  ! AX140884 Sequence 374 from Pat
gb_pat:AX200744       + 2024.00 2804.04  8.0e-148  2000  ! AX200744 Sequence 374 from Pat
gb_pat:AX267400       + 2024.00 2804.04  8.0e-148  2000  ! AX267400 Sequence 374 from Pat
gb_pat:AX282957       + 2024.00 2804.04  8.0e-148  2000  ! AX282957 Sequence 6 from Paten
gb_pat:AX316965       + 2024.00 2804.04  8.0e-148  2000  ! AX316965 Sequence 302 from Pat
gb_pat:AX106594       + 2024.00 2803.85  8.2e-148  2040  ! AX106594 Sequence 375 from Pat
gb_pat:AX140885       + 2024.00 2803.85  8.2e-148  2040  ! AX140885 Sequence 375 from Pat
gb_pat:AX200745       + 2024.00 2803.85  8.2e-148  2040  ! AX200745 Sequence 375 from Pat
gb_pat:AX267401       + 2024.00 2803.85  8.2e-148  2040  ! AX267401 Sequence 375 from Pat
gb_pat:AX282958       + 2024.00 2803.85  8.2e-148  2040  ! AX282958 Sequence 7 from Paten
gb_pat:AX316966       + 2024.00 2803.85  8.2e-148  2040  ! AX316966 Sequence 303 from Pat
gb_pat:AX106588       + 1920.50 2660.74  7.7e-140  1853  ! AX106588 Sequence 369 from Pat
gb_pat:AX140879       + 1920.50 2660.74  7.7e-140  1853  ! AX140879 Sequence 369 from Pat
gb_pat:AX200739       + 1920.50 2660.74  7.7e-140  1853  ! AX200739 Sequence 369 from Pat
gb_pat:AX267395       + 1920.50 2660.74  7.7e-140  1853  ! AX267395 Sequence 369 from Pat
gb_pat:AX316958       + 1920.50 2660.74  7.7e-140  1853  ! AX316958 Sequence 295 from Pat
gb_pat:AX316992       + 1876.00 2603.17  1.2e-136  1155  ! AX316992 Sequence 329 from Pat
gb_pat:AX316988       + 1857.00 2576.73  3.7e-135  1155  ! AX316988 Sequence 325 from Pat
gb_pat:AX316993       + 1857.00 2576.73  3.7e-135  1155  ! AX316993 Sequence 330 from Pat
gb_pat:AX141040       + 1495.00 2068.74  7.2e-107  1852  ! AX141040 Sequence 530 from Pat
gb_pat:AX200900       + 1495.00 2068.74  7.2e-107  1852  ! AX200900 Sequence 530 from Pat
gb_pat:AX267556       + 1495.00 2068.74  7.2e-107  1852  ! AX267556 Sequence 530 from Pat
gb_pat:AX282954       + 1495.00 2068.74  7.2e-107  1852  ! AX282954 Sequence 3 from Paten
gb_pat:AX316976       + 1495.00 2068.74  7.2e-107  1852  ! AX316976 Sequence 313 from Pat
gb_pat:AX106587       + 1469.00 2034.43  5.9e-105  1512  ! AX106587 Sequence 368 from Pat
gb_pat:AX140878       + 1469.00 2034.43  5.9e-105  1512  ! AX140878 Sequence 368 from Pat
gb_pat:AX200738       + 1469.00 2034.43  5.9e-105  1512  ! AX200738 Sequence 368 from Pat
gb_pat:AX267394       + 1469.00 2034.43  5.9e-105  1512  ! AX267394 Sequence 368 from Pat
gb_pat:AX316957       + 1469.00 2034.43  5.9e-105  1512  ! AX316957 Sequence 294 from Pat
gb_pat:AR148112       - 1468.00 2031.18  8.9e-105  1851  ! AR148112 Sequence 291 from pat
gb_pat:AX106585       - 1468.00 2031.18  8.9e-105  1851  ! AX106585 Sequence 366 from Pat
gb_pat:AX140876       - 1468.00 2031.18  8.9e-105  1851  ! AX140876 Sequence 366 from
gb_pat:AX200736       - 1468.00 2031.18  8.9e-105  1851  ! AX200736 Sequence 366 from
gb_pat:AX267392       - 1468.00 2031.18  8.9e-105  1851  ! AX267392 Sequence 366 from
gb_pat:AX282952       - 1468.00 2031.18  8.9e-105  1851  ! AX282952 Sequence 1 from Pa
gb_pat:AX316954       - 1468.00 2031.18  8.9e-105  1851  ! AX316954 Sequence 291 from

seq_name: gb_pat:AX106592

seq_documentation_block:
LOCUS       AX106592               1155 bp    DNA     linear   PAT 30-APR-2001
DEFINITION  Sequence 373 from Patent WO0125272.
ACCESSION   AX106592
VERSION     AX106592.1  GI:13922263
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1155)
  AUTHORS   Xu,J., Skeiky,Y.A., Reed,S.G. and Cheever,M.A.
  TITLE     Compositions and methods for therapy and diagnosis of prostate
            cancer
  JOURNAL   Patent: WO 0125272-A 373 12-APR-2001;
            CORIXA CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1. .1155
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      346 a    253 c    297 g    259 t
ORIGIN


alignment_scores:
            Quality: 2064.00             Length:    384
              Ratio:  5.375                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-810-936-304 x AX106592   ..

 Align seg 1/1  to: AX106592  from: 1  to: 1155

       1 MetValValGluValAspSerMetProAlaAlaSerSerValLysLysPr 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 ATGGTGGTTGAGGTTGATTCCATGCCGGCTGCCTCTTCTGTGAAGAAGCC 50

      17 oPheGlyLeuArgSerLysMetGlyLysTrpCysCysArgCysPheProC 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 ATTTGGTCTCAGGAGCAAGATGGGCAAGTGGTGCTGCCGTTGCTTCCCCT 100

      34 ysCysArgGluSerGlyLysSerAsnValGlyThrSerGlyAspHisAsp 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 GCTGCAGGGAGAGCGGCAAGAGCAACGTGGGCACTTCTGGAGACCACGAC 150

      51 AspSerAlaMetLysThrLeuArgSerLysMetGlyLysTrpCysArgHi 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     151 GACTCTGCTATGAAGACACTCAGGAGCAAGATGGGCAAGTGGTGCCGCCA 200

      67 sCysPheProCysCysArgGlySerGlyLysSerAsnValGlyAlaSerG 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     201 CTGCTTCCCCTGCTGCAGGGGGAGTGGCAAGAGCAACGTGGGCGCTTCTG 250

      84 lyAspHisAspAspSerAlaMetLysThrLeuArgAsnLysMetGlyLys 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     251 GAGACCACGACGACTCTGCTATGAAGACACTCAGGAACAAGATGGGCAAG 300

     101 TrpCysCysHisCysPheProCysCysArgGlySerGlyLysSerLysVa 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     301 TGGTGCTGCCACTGCTTCCCCTGCTGCAGGGGGAGCGGCAAGAGCAAGGT 350

     117 lGlyAlaTrpGlyAspTyrAspAspSerAlaPheMetGluProArgTyrH 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     351 GGGCGCTTGGGGAGACTACGATGACAGTGCCTTCATGGAGCCCAGGTACC 400
```

```
  134 isValArgGlyGluAspLeuAspLysLeuHisArgAlaAlaTrpTrpGly 150
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  401 ACGTCCGTGGAGAAGATCTGGACAAGCTCCACAGAGCTGCCTGGTGGGGT 450

  151 LysValProArgLysAspLeuIleValMetLeuArgAspThrAspValAs 167
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  451 AAAGTCCCCAGAAAGGATCTCATCGTCATGCTCAGGGACACTGACGTGAA 500

  167 nLysLysAspLysGlnLysArgThrAlaLeuHisLeuAlaSerAlaAsnG 184
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  501 CAAGAAGGACAAGCAAAAGAGGACTGCTCTACATCTGGCCTCTGCCAATG 550

  184 lyAsnSerGluValValLysLeuLeuLeuAspArgArgCysGlnLeuAsn 200
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  551 GGAATTCAGAAGTAGTAAAACTCCTGCTGGACAGACGATGTCAACTTAAT 600

  201 ValLeuAspAsnLysLysArgThrAlaLeuIleLysAlaValGlnCysGl 217
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  601 GTCCTTGACAACAAAAAGAGGACAGCTCTGATAAAGGCCGTACAATGCCA 650

  217 nGluAspGluCysAlaLeuMetLeuLeuGluHisGlyThrAspProAsnI 234
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  651 GGAAGATGAATGTGCGTTAATGTTGCTGGAACATGGCACTGATCCAAATA 700

  234 leProAspGluTyrGlyAsnThrThrLeuHisTyrAlaIleTyrAsnGlu 250
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  701 TTCCAGATGAGTATGGAAATACCACTCTGCACTACGCTATCTATAATGAA 750

  251 AspLysLeuMetAlaLysAlaLeuLeuLeuTyrGlyAlaAspIleGluSe 267
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  751 GATAAATTAATGGCCAAAGCACTGCTCTTATATGGTGCTGATATCGAATC 800

  267 rLysAsnLysHisGlyLeuThrProLeuLeuLeuGlyValHisGluGlnL 284
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  801 AAAAAACAAGCATGGCCTCACACCACTGTTACTTGGTGTACATGAGCAAA 850

  284 ysGlnGlnValValLysPheLeuIleLysLysLysAlaAsnLeuAsnAla 300
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  851 AACAGCAAGTCGTGAAATTTTTAATCAAGAAAAAAGCGAATTTAAATGCA 900

  301 LeuAspArgTyrGlyArgThrAlaLeuIleLeuAlaValCysCysGlySe 317
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  901 CTGGATAGATATGGAAGGACTGCTCTCATACTTGCTGTATGTTGTGGATC 950

  317 rAlaSerIleValSerLeuLeuLeuGluGlnAsnIleAspValSerSerG 334
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  951 AGCAAGTATAGTCAGCCTTCTACTTGAGCAAAATATTGATGTATCTTCTC 1000

  334 lnAspLeuSerGlyGlnThrAlaArgGluTyrAlaValSerSerHisHis 350
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1001 AAGATCTATCTGGACAGACGGCCAGAGAGTATGCTGTTTCTAGTCATCAT 1050

  351 HisValIleCysGlnLeuLeuSerAspTyrLysGluLysGlnMetLeuLy 367
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1051 CATGTAATTTGCCAGTTACTTTCTGACTACAAAGAAAAACAGATGCTAAA 1100

  367 sIleSerSerGluAsnSerAsnProGluAsnValSerArgThrArgAsnL 384
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1101 AATCTCTTCTGAAAACAGCAATCCAGAAAATGTCTCAAGAACCAGAAATA 1150

  384 ys 384
      ||
 1151 AA 1152

seq_name: gb_pat:AX140883

seq_documentation_block:
LOCUS       AX140883               1155 bp    DNA     linear   PAT 31-MAY-2001
DEFINITION  Sequence 373 from Patent WO0134802.
ACCESSION   AX140883
VERSION     AX140883.1  GI:14280986
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1155)
  AUTHORS   Xu,J., Dillon,D.C., Mitcham,J.L., Harlocker,S.L., Jiang,Y.,
            Reed,S.G., Kalos,M.D., Retter,M.W., Stolk,J.A., Day,C.H.,
            Skeiky,Y.A. and Wang,A.
  TITLE     Compositions and methods for the therapy and diagnosis of prostate
            cancer
  JOURNAL   Patent: WO 0134802-A 373 17-MAY-2001;
            CORIXA CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1. .1155
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      346 a    253 c    297 g    259 t
ORIGIN


alignment_scores:
            Quality: 2064.00              Length:    384
              Ratio:  5.375                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-810-936-304 x AX140883   ..

 Align seg 1/1  to: AX140883  from: 1  to: 1155

    1 MetValValGluValAspSerMetProAlaAlaSerSerValLysLysPr 17
      ||||||||||||||||||||||||||||||||||||||||||||||||||
    1 ATGGTGGTTGAGGTTGATTCCATGCCGGCTGCCTCTTCTGTGAAGAAGCC 50

   17 oPheGlyLeuArgSerLysMetGlyLysTrpCysCysArgCysPheProC 34
      ||||||||||||||||||||||||||||||||||||||||||||||||||
   51 ATTTGGTCTCAGGAGCAAGATGGGCAAGTGGTGCTGCCGTTGCTTCCCCT 100

   34 ysCysArgGluSerGlyLysSerAsnValGlyThrSerGlyAspHisAsp 50
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  101 GCTGCAGGGAGAGCGGCAAGAGCAACGTGGGCACTTCTGGAGACCACGAC 150

   51 AspSerAlaMetLysThrLeuArgSerLysMetGlyLysTrpCysArgHi 67
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  151 GACTCTGCTATGAAGACACTCAGGAGCAAGATGGGCAAGTGGTGCCGCCA 200

   67 sCysPheProCysCysArgGlySerGlyLysSerAsnValGlyAlaSerG 84
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  201 CTGCTTCCCCTGCTGCAGGGGGAGTGGCAAGAGCAACGTGGGCGCTTCTG 250

   84 lyAspHisAspAspSerAlaMetLysThrLeuArgAsnLysMetGlyLys 100
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  251 GAGACCACGACGACTCTGCTATGAAGACACTCAGGAACAAGATGGGCAAG 300

  101 TrpCysCysHisCysPheProCysCysArgGlySerGlyLysSerLysVa 117
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  301 TGGTGCTGCCACTGCTTCCCCTGCTGCAGGGGGAGCGGCAAGAGCAAGGT 350

  117 lGlyAlaTrpGlyAspTyrAspAspSerAlaPheMetGluProArgTyrH 134
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  351 GGGCGCTTGGGGAGACTACGATGACAGTGCCTTCATGGAGCCCAGGTACC 400

  134 isValArgGlyGluAspLeuAspLysLeuHisArgAlaAlaTrpTrpGly 150
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  401 ACGTCCGTGGAGAAGATCTGGACAAGCTCCACAGAGCTGCCTGGTGGGGT 450

  151 LysValProArgLysAspLeuIleValMetLeuArgAspThrAspValAs 167
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  451 AAAGTCCCCAGAAAGGATCTCATCGTCATGCTCAGGGACACTGACGTGAA 500

  167 nLysLysAspLysGlnLysArgThrAlaLeuHisLeuAlaSerAlaAsnG 184
      ||||||||||||||||||||||||||||||||||||||||||||||||||
  501 CAAGAAGGACAAGCAAAAGAGGACTGCTCTACATCTGGCCTCTGCCAATG 550
```

```
    184 lyAsnSerGluValValLysLeuLeuLeuAspArgArgCysGlnLeuAsn 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    551 GGAATTCAGAAGTAGTAAAACTCCTGCTGGACAGACGATGTCAACTTAAT 600

    201 ValLeuAspAsnLysLysArgThrAlaLeuIleLysAlaValGlnCysGl 217
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    601 GTCCTTGACAACAAAAAGAGGACAGCTCTGATAAAGGCCGTACAATGCCA 650

    217 nGluAspGluCysAlaLeuMetLeuLeuGluHisGlyThrAspProAsnI 234
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    651 GGAAGATGAATGTGCGTTAATGTTGCTGGAACATGGCACTGATCCAAATA 700

    234 leProAspGluTyrGlyAsnThrThrLeuHisTyrAlaIleTyrAsnGlu 250
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    701 TTCCAGATGAGTATGGAAATACCACTCTGCACTACGCTATCTATAATGAA 750

    251 AspLysLeuMetAlaLysAlaLeuLeuLeuTyrGlyAlaAspIleGluSe 267
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    751 GATAAATTAATGGCCAAAGCACTGCTCTTATATGGTGCTGATATCGAATC 800

    267 rLysAsnLysHisGlyLeuThrProLeuLeuLeuGlyValHisGluGlnL 284
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    801 AAAAAACAAGCATGGCCTCACACCACTGTTACTTGGTGTACATGAGCAAA 850

    284 ysGlnGlnValValLysPheLeuIleLysLysLysAlaAsnLeuAsnAla 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    851 AACAGCAAGTCGTGAAATTTTTAATCAAGAAAAAAGCGAATTTAAATGCA 900

    301 LeuAspArgTyrGlyArgThrAlaLeuIleLeuAlaValCysCysGlySe 317
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    901 CTGGATAGATATGGAAGGACTGCTCTCATACTTGCTGTATGTTGTGGATC 950

    317 rAlaSerIleValSerLeuLeuLeuGluGlnAsnIleAspValSerSerG 334
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    951 AGCAAGTATAGTCAGCCTTCTACTTGAGCAAAATATTGATGTATCTTCTC 1000

    334 lnAspLeuSerGlyGlnThrAlaArgGluTyrAlaValSerSerHisHis 350
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
   1001 AAGATCTATCTGGACAGACGGCCAGAGAGTATGCTGTTTCTAGTCATCAT 1050

    351 HisValIleCysGlnLeuLeuSerAspTyrLysGluLysGlnMetLeuLy 367
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
   1051 CATGTAATTTGCCAGTTACTTTCTGACTACAAAGAAAAACAGATGCTAAA 1100

    367 sIleSerSerGluAsnSerAsnProGluAsnValSerArgThrArgAsnL 384
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
   1101 AATCTCTTCTGAAAACAGCAATCCAGAAAATGTCTCAAGAACCAGAAATA 1150

    384 ys 384
        ||
   1151 AA 1152

seq_name: gb_pat:AX200743

seq_documentation_block:
LOCUS       AX200743               1155 bp    DNA     linear   PAT 29-AUG-2001
DEFINITION  Sequence 373 from Patent WO0151633.
ACCESSION   AX200743
VERSION     AX200743.1  GI:15390636
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1155)
  AUTHORS   Xu,J., Dillon,D.C., Mitcham,J.L., Harlocker,S.L., Jiang,Y.,
            Reed,S.G., Kalos,M.D., Fanger,G.R., Day,C.H., Retter,M.W.,
            Stolk,J.A., Skeiky,Y.A., Wang,A. and Meagher,M.J.
  TITLE     Compositions and methods for the therapy and diagnosis of prostate
            cancer
  JOURNAL   Patent: WO 0151633-A 373 19-JUL-2001;
            CORIXA CORPORATION (US)

FEATURES             Location/Qualifiers
     source          1. .1155
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      346 a    253 c    297 g    259 t
ORIGIN


alignment_scores:
            Quality: 2064.00              Length:    384
              Ratio:  5.375                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-810-936-304 x AX200743   ..

 Align seg 1/1  to: AX200743  from: 1  to: 1155

      1 MetValValGluValAspSerMetProAlaAlaSerSerValLysLysPr 17
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
      1 ATGGTGGTTGAGGTTGATTCCATGCCGGCTGCCTCTTCTGTGAAGAAGCC 50

     17 oPheGlyLeuArgSerLysMetGlyLysTrpCysCysArgCysPheProC 34
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
     51 ATTTGGTCTCAGGAGCAAGATGGGCAAGTGGTGCTGCCGTTGCTTCCCCT 100

     34 ysCysArgGluSerGlyLysSerAsnValGlyThrSerGlyAspHisAsp 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    101 GCTGCAGGGAGAGCGGCAAGAGCAACGTGGGCACTTCTGGAGACCACGAC 150

     51 AspSerAlaMetLysThrLeuArgSerLysMetGlyLysTrpCysArgHi 67
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    151 GACTCTGCTATGAAGACACTCAGGAGCAAGATGGGCAAGTGGTGCCGCCA 200

     67 sCysPheProCysCysArgGlySerGlyLysSerAsnValGlyAlaSerG 84
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    201 CTGCTTCCCCTGCTGCAGGGGGAGTGGCAAGAGCAACGTGGGCGCTTCTG 250

     84 lyAspHisAspAspSerAlaMetLysThrLeuArgAsnLysMetGlyLys 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    251 GAGACCACGACGACTCTGCTATGAAGACACTCAGGAACAAGATGGGCAAG 300

    101 TrpCysCysHisCysPheProCysCysArgGlySerGlyLysSerLysVa 117
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    301 TGGTGCTGCCACTGCTTCCCCTGCTGCAGGGGGAGCGGCAAGAGCAAGGT 350

    117 lGlyAlaTrpGlyAspTyrAspAspSerAlaPheMetGluProArgTyrH 134
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    351 GGGCGCTTGGGGAGACTACGATGACAGTGCCTTCATGGAGCCCAGGTACC 400

    134 isValArgGlyGluAspLeuAspLysLeuHisArgAlaAlaTrpTrpGly 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    401 ACGTCCGTGGAGAAGATCTGGACAAGCTCCACAGAGCTGCCTGGTGGGGT 450

    151 LysValProArgLysAspLeuIleValMetLeuArgAspThrAspValAs 167
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    451 AAAGTCCCCAGAAAGGATCTCATCGTCATGCTCAGGGACACTGACGTGAA 500

    167 nLysLysAspLysGlnLysArgThrAlaLeuHisLeuAlaSerAlaAsnG 184
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    501 CAAGAAGGACAAGCAAAAGAGGACTGCTCTACATCTGGCCTCTGCCAATG 550

    184 lyAsnSerGluValValLysLeuLeuLeuAspArgArgCysGlnLeuAsn 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    551 GGAATTCAGAAGTAGTAAAACTCCTGCTGGACAGACGATGTCAACTTAAT 600

    201 ValLeuAspAsnLysLysArgThrAlaLeuIleLysAlaValGlnCysGl 217
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    601 GTCCTTGACAACAAAAAGAGGACAGCTCTGATAAAGGCCGTACAATGCCA 650

    217 nGluAspGluCysAlaLeuMetLeuLeuGluHisGlyThrAspProAsnI 234
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    651 GGAAGATGAATGTGCGTTAATGTTGCTGGAACATGGCACTGATCCAAATA 700
```

```
     234 leProAspGluTyrGlyAsnThrThrLeuHisTyrAlaIleTyrAsnGlu 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     701 TTCCAGATGAGTATGGAAATACCACTCTGCACTACGCTATCTATAATGAA 750

     251 AspLysLeuMetAlaLysAlaLeuLeuLeuTyrGlyAlaAspIleGluSe 267
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     751 GATAAATTAATGGCCAAAGCACTGCTCTTATATGGTGCTGATATCGAATC 800

     267 rLysAsnLysHisGlyLeuThrProLeuLeuLeuGlyValHisGluGlnL 284
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     801 AAAAAACAAGCATGGCCTCACACCACTGTTACTTGGTGTACATGAGCAAA 850

     284 ysGlnGlnValValLysPheLeuIleLysLysLysAlaAsnLeuAsnAla 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     851 AACAGCAAGTCGTGAAATTTTTAATCAAGAAAAAAGCGAATTTAAATGCA 900

     301 LeuAspArgTyrGlyArgThrAlaLeuIleLeuAlaValCysCysGlySe 317
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     901 CTGGATAGATATGGAAGGACTGCTCTCATACTTGCTGTATGTTGTGGATC 950

     317 rAlaSerIleValSerLeuLeuLeuGluGlnAsnIleAspValSerSerG 334
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     951 AGCAAGTATAGTCAGCCTTCTACTTGAGCAAAATATTGATGTATCTTCTC 1000

     334 lnAspLeuSerGlyGlnThrAlaArgGluTyrAlaValSerSerHisHis 350
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1001 AAGATCTATCTGGACAGACGGCCAGAGAGTATGCTGTTTCTAGTCATCAT 1050

     351 HisValIleCysGlnLeuLeuSerAspTyrLysGluLysGlnMetLeuLy 367
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1051 CATGTAATTTGCCAGTTACTTTCTGACTACAAAGAAAAACAGATGCTAAA 1100

     367 sIleSerSerGluAsnSerAsnProGluAsnValSerArgThrArgAsnL 384
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1101 AATCTCTTCTGAAAACAGCAATCCAGAAAATGTCTCAAGAACCAGAAATA 1150

     384 ys 384
         ||
    1151 AA 1152

seq_name: gb_pat:AX267399

seq_documentation_block:
LOCUS       AX267399               1155 bp    DNA     linear   PAT 26-OCT-2001
DEFINITION  Sequence 373 from Patent WO0173032.
ACCESSION   AX267399
VERSION     AX267399.1  GI:16516169
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Xu,J., Dillon,D.C., Mitcham,J.L., Harlocker,S.L., Jiang,Y.,
            Kalos,M.D., Fanger,G.R., Retter,M.W., Stolk,J.A., Day,C.H.,
            Vedvick,T.S., Carter,D., Li,S.X., Wang,A., Skeiky,Y.A., Hepler,W.T.
            and Henderson,R.A.
  TITLE     Compositions and methods for the therapy and diagnosis of prostate
            cancer
  JOURNAL   Patent: WO 0173032-A 373 04-OCT-2001;
            CORIXA CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1. .1155
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      346 a    253 c    297 g    259 t
ORIGIN


alignment_scores:
            Quality: 2064.00             Length:    384
              Ratio:  5.375                Gaps:      0

 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-810-936-304 x AX267399   ..

 Align seg 1/1  to: AX267399  from: 1  to: 1155

       1 MetValValGluValAspSerMetProAlaAlaSerSerValLysLysPr 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 ATGGTGGTTGAGGTTGATTCCATGCCGGCTGCCTCTTCTGTGAAGAAGCC 50

      17 oPheGlyLeuArgSerLysMetGlyLysTrpCysCysArgCysPheProC 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 ATTTGGTCTCAGGAGCAAGATGGGCAAGTGGTGCTGCCGTTGCTTCCCCT 100

      34 ysCysArgGluSerGlyLysSerAsnValGlyThrSerGlyAspHisAsp 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 GCTGCAGGGAGAGCGGCAAGAGCAACGTGGGCACTTCTGGAGACCACGAC 150

      51 AspSerAlaMetLysThrLeuArgSerLysMetGlyLysTrpCysArgHi 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     151 GACTCTGCTATGAAGACACTCAGGAGCAAGATGGGCAAGTGGTGCCGCCA 200

      67 sCysProCysCysArgGlySerGlyLysSerAsnValGlyAlaSerG 84
         ||||||||||||||||||||||||||||||||||||||||||||||||
     201 CTGCTTCCCCTGCTGCAGGGGGAGTGGCAAGAGCAACGTGGGCGCTTCTG 250

      84 lyAspHisAspAspSerAlaMetLysThrLeuArgAsnLysMetGlyLys 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     251 GAGACCACGACGACTCTGCTATGAAGACACTCAGGAACAAGATGGGCAAG 300

     101 TrpCysCysHisCysProCysCysArgGlySerGlyLysSerLysVa 117
         ||||||||||||||||||||||||||||||||||||||||||||||||
     301 TGGTGCTGCCACTGCTTCCCCTGCTGCAGGGGGAGCGGCAAGAGCAAGGT 350

     117 lGlyAlaTrpGlyAspTyrAspAspSerAlaPheMetGluProArgTyrH 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     351 GGGCGCTTGGGGAGACTACGATGACAGTGCCTTCATGGAGCCCAGGTACC 400

     134 isValArgGlyGluAspLeuAspLysLeuHisArgAlaAlaTrpTrpGly 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     401 ACGTCCGTGGAGAAGATCTGGACAAGCTCCACAGAGCTGCCTGGTGGGGT 450

     151 LysValProArgLysAspLeuIleValMetLeuArgAspThrAspValAs 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     451 AAAGTCCCCAGAAAGGATCTCATCGTCATGCTCAGGGACACTGACGTGAA 500

     167 nLysLysAspLysGlnLysArgThrAlaLeuHisLeuAlaSerAlaAsnG 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     501 CAAGAAGGACAAGCAAAAGAGGACTGCTCTACATCTGGCCTCTGCCAATG 550

     184 lyAsnSerGluValValLysLeuLeuLeuAspArgArgCysGlnLeuAsn 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     551 GGAATTCAGAAGTAGTAAAACTCCTGCTGGACAGACGATGTCAACTTAAT 600

     201 ValLeuAspAsnLysLysArgThrAlaLeuIleLysAlaValGlnCysGl 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     601 GTCCTTGACAACAAAAAGAGGACAGCTCTGATAAAGGCCGTACAATGCCA 650

     217 nGluAspGluCysAlaLeuMetLeuLeuGluHisGlyThrAspProAsnI 234
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     651 GGAAGATGAATGTGCGTTAATGTTGCTGGAACATGGCACTGATCCAAATA 700

     234 leProAspGluTyrGlyAsnThrThrLeuHisTyrAlaIleTyrAsnGlu 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     701 TTCCAGATGAGTATGGAAATACCACTCTGCACTACGCTATCTATAATGAA 750

     251 AspLysLeuMetAlaLysAlaLeuLeuLeuTyrGlyAlaAspIleGluSe 267
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     751 GATAAATTAATGGCCAAAGCACTGCTCTTATATGGTGCTGATATCGAATC 800

     267 rLysAsnLysHisGlyLeuThrProLeuLeuLeuGlyValHisGluGlnL 284
         ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
     801 AAAAAACAAGCATGGCCTCACACCACTGTTACTTGGTGTACATGAGCAAA 850

     284 ysGlnGlnValValLysPheLeuIleLysLysLysAlaAsnLeuAsnAla 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     851 AACAGCAAGTCGTGAAATTTTTAATCAAGAAAAAAGCGAATTTAAATGCA 900

     301 LeuAspArgTyrGlyArgThrAlaLeuIleLeuAlaValCysCysGlySe 317
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     901 CTGGATAGATATGGAAGGACTGCTCTCATACTTGCTGTATGTTGTGGATC 950

     317 rAlaSerIleValSerLeuLeuLeuGluGlnAsnIleAspValSerSerG 334
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     951 AGCAAGTATAGTCAGCCTTCTACTTGAGCAAAATATTGATGTATCTTCTC 1000

     334 lnAspLeuSerGlyGlnThrAlaArgGluTyrAlaValSerSerHisHis 350
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    1001 AAGATCTATCTGGACAGACGGCCAGAGAGTATGCTGTTTCTAGTCATCAT 1050

     351 HisValIleCysGlnLeuLeuSerAspTyrLysGluLysGlnMetLeuLy 367
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    1051 CATGTAATTTGCCAGTTACTTTCTGACTACAAAGAAAAACAGATGCTAAA 1100

     367 sIleSerSerGluAsnSerAsnProGluAsnValSerArgThrArgAsnL 384
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    1101 AATCTCTTCTGAAAACAGCAATCCAGAAAATGTCTCAAGAACCAGAAATA 1150

     384 ys 384
         ||
    1151 AA 1152

seq_name: gb_pat:AX282956

seq_documentation_block:
LOCUS       AX282956                1155 bp    DNA     linear   PAT 02-NOV-2001
DEFINITION  Sequence 5 from Patent WO0175171.
ACCESSION   AX282956
VERSION     AX282956.1  GI:16609896
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Houghton,R.L., Dillon,D.C., Molesh,D.A., Xu,J., Zehentner,B. and
            Persing,D.H.
  TITLE     Methods, compositions and kits for the detection and monitoring of
            breast cancer
  JOURNAL   Patent: WO 0175171-A 5 11-OCT-2001;
            CORIXA CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1. .1155
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      346 a    253 c    297 g    259 t
ORIGIN


alignment_scores:
            Quality: 2064.00             Length:    384
              Ratio:  5.375                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-810-936-304 x AX282956   ..

 Align seg 1/1  to: AX282956  from: 1  to: 1155

       1 MetValValGluValAspSerMetProAlaAlaSerSerValLysLysPr 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 ATGGTGGTTGAGGTTGATTCCATGCCGGCTGCCTCTTCTGTGAAGAAGCC 50

      17 oPheGlyLeuArgSerLysMetGlyLysTrpCysCysArgCysPheProC 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||

      51 ATTTGGTCTCAGGAGCAAGATGGGCAAGTGGTGCTGCCGTTGCTTCCCCT 100

      34 ysCysArgGluSerGlyLysSerAsnValGlyThrSerGlyAspHisAsp 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 GCTGCAGGGAGAGCGGCAAGAGCAACGTGGGCACTTCTGGAGACCACGAC 150

      51 AspSerAlaMetLysThrLeuArgSerLysMetGlyLysTrpCysArgHi 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     151 GACTCTGCTATGAAGACACTCAGGAGCAAGATGGGCAAGTGGTGCCGCCA 200

      67 sCysProCysCysArgGlySerGlyLysSerAsnValGlyAlaSerG 84
         ||||||||||||||||||||||||||||||||||||||||||||||||
     201 CTGCTTCCCCTGCTGCAGGGGAGTGGCAAGAGCAACGTGGGCGCTTCTG 250

      84 lyAspHisAspAspSerAlaMetLysThrLeuArgAsnLysMetGlyLys 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     251 GAGACCACGACGACTCTGCTATGAAGACACTCAGGAACAAGATGGGCAAG 300

     101 TrpCysCysHisCysProCysCysArgGlySerGlyLysSerLysVa 117
         ||||||||||||||||||||||||||||||||||||||||||||||||
     301 TGGTGCTGCCACTGCTTCCCCTGCTGCAGGGGGAGCGGCAAGAGCAAGGT 350

     117 lGlyAlaTrpGlyAspTyrAspAspSerAlaPheMetGluProArgTyrH 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     351 GGGCGCTTGGGGAGACTACGATGACAGTGCCTTCATGGAGCCCAGGTACC 400

     134 isValArgGlyGluAspLeuAspLysLeuHisArgAlaAlaTrpTrpGly 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     401 ACGTCCGTGGAGAAGATCTGGACAAGCTCCACAGAGCTGCCTGGTGGGGT 450

     151 LysValProArgLysAspLeuIleValMetLeuArgAspThrAspValAs 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     451 AAAGTCCCCAGAAAGGATCTCATCGTCATGCTCAGGGACACTGACGTGAA 500

     167 nLysLysAspLysGlnLysArgThrAlaLeuHisLeuAlaSerAlaAsnG 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     501 CAAGAAGGACAAGCAAAAGAGGACTGCTCTACATCTGGCCTCTGCCAATG 550

     184 lyAsnSerGluValValLysLeuLeuLeuAspArgArgCysGlnLeuAsn 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     551 GGAATTCAGAAGTAGTAAAACTCCTGCTGGACAGACGATGTCAACTTAAT 600

     201 ValLeuAspAsnLysLysArgThrAlaLeuIleLysAlaValGlnCysGl 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     601 GTCCTTGACAACAAAAAGAGGACAGCTCTGATAAAGGCCGTACAATGCCA 650

     217 nGluAspGluCysAlaLeuMetLeuLeuGluHisGlyThrAspProAsnI 234
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     651 GGAAGATGAATGTGCGTTAATGTTGCTGGAACATGGCACTGATCCAAATA 700

     234 leProAspGluTyrGlyAsnThrThrLeuHisTyrAlaIleTyrAsnGlu 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     701 TTCCAGATGAGTATGGAAATACCACTCTGCACTACGCTATCTATAATGAA 750

     251 AspLysLeuMetAlaLysAlaLeuLeuLeuTyrGlyAlaAspIleGluSe 267
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     751 GATAAATTAATGGCCAAAGCACTGCTCTTATATGGTGCTGATATCGAATC 800

     267 rLysAsnLysHisGlyLeuThrProLeuLeuLeuGlyValHisGluGlnL 284
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     801 AAAAAACAAGCATGGCCTCACACCACTGTTACTTGGTGTACATGAGCAAA 850

     284 ysGlnGlnValValLysPheLeuIleLysLysLysAlaAsnLeuAsnAla 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     851 AACAGCAAGTCGTGAAATTTTTAATCAAGAAAAAAGCGAATTTAAATGCA 900

     301 LeuAspArgTyrGlyArgThrAlaLeuIleLeuAlaValCysCysGlySe 317
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     901 CTGGATAGATATGGAAGGACTGCTCTCATACTTGCTGTATGTTGTGGATC 950

     317 rAlaSerIleValSerLeuLeuLeuGluGlnAsnIleAspValSerSerG 334
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     951 AGCAAGTATAGTCAGCCTTCTACTTGAGCAAAATATTGATGTATCTTCTC 1000
```

```
     334 lnAspLeuSerGlyGlnThrAlaArgGluTyrAlaValSerSerHisHis 350
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1001 AAGATCTATCTGGACAGACGGCCAGAGAGTATGCTGTTTCTAGTCATCAT 1050

     351 HisValIleCysGlnLeuLeuSerAspTyrLysGluLysGlnMetLeuLy 367
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1051 CATGTAATTTGCCAGTTACTTTCTGACTACAAAGAAAAACAGATGCTAAA 1100

     367 sIleSerSerGluAsnSerAsnProGluAsnValSerArgThrArgAsnL 384
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1101 AATCTCTTCTGAAAACAGCAATCCAGAAAATGTCTCAAGAACCAGAAATA 1150

     384 ys 384
         ||
    1151 AA 1152

seq_name: gb_pat:AX316964

seq_documentation_block:
LOCUS       AX316964               1155 bp    DNA     linear   PAT 14-DEC-2001
DEFINITION  Sequence 301 from Patent WO0190152.
ACCESSION   AX316964
VERSION     AX316964.1  GI:17900043
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Frudakis,T.N., Reed,S.G., Smith,J.M., Misher,L.E., Dillon,D.C.,
            Retter,M.W., Wang,A., Skeiky,Y.A., Harlocker,S.L. and Day,C.H.
  TITLE     Compositions and methods for the therapy and diagnosis of breast
            cancer
  JOURNAL   Patent: WO 0190152-A 301 29-NOV-2001;
            CORIXA CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1. .1155
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      346 a    253 c    297 g    259 t
ORIGIN


alignment_scores:
            Quality: 2064.00            Length:    384
              Ratio:  5.375               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-810-936-304 x AX316964   ..

 Align seg 1/1  to: AX316964  from: 1  to: 1155

       1 MetValValGluValAspSerMetProAlaAlaSerSerValLysLysPr 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 ATGGTGGTTGAGGTTGATTCCATGCCGGCTGCCTCTTCTGTGAAGAAGCC 50

      17 oPheGlyLeuArgSerLysMetGlyLysTrpCysCysArgCysPheProC 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 ATTTGGTCTCAGGAGCAAGATGGGCAAGTGGTGCTGCCGTTGCTTCCCCT 100

      34 ysCysArgGluSerGlyLysSerAsnValGlyThrSerGlyAspHisAsp 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 GCTGCAGGGAGAGCGGCAAGAGCAACGTGGGCACTTCTGGAGACCACGAC 150

      51 AspSerAlaMetLysThrLeuArgSerLysMetGlyLysTrpCysArgHi 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     151 GACTCTGCTATGAAGACACTCAGGAGCAAGATGGGCAAGTGGTGCCGCCA 200

      67 sCysPheProCysCysArgGlySerGlyLysSerAsnValGlyAlaSerG 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     201 CTGCTTCCCCTGCTGCAGGGGGAGTGGCAAGAGCAACGTGGGCGCTTCTG 250

      84 lyAspHisAspAspSerAlaMetLysThrLeuArgAsnLysMetGlyLys 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     251 GAGACCACGACGACTCTGCTATGAAGACACTCAGGAACAAGATGGGCAAG 300

     101 TrpCysCysHisCysPheProCysCysArgGlySerGlyLysSerLysVa 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     301 TGGTGCTGCCACTGCTTCCCCTGCTGCAGGGGGAGCGGCAAGAGCAAGGT 350

     117 lGlyAlaTrpGlyAspTyrAspAspSerAlaPheMetGluProArgTyrH 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     351 GGGCGCTTGGGGAGACTACGATGACAGTGCCTTCATGGAGCCCAGGTACC 400

     134 isValArgGlyGluAspLeuAspLysLeuHisArgAlaAlaTrpTrpGly 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     401 ACGTCCGTGGAGAAGATCTGGACAAGCTCCACAGAGCTGCCTGGTGGGGT 450

     151 LysValProArgLysAspLeuIleValMetLeuArgAspThrAspValAs 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     451 AAAGTCCCCAGAAAGGATCTCATCGTCATGCTCAGGGACACTGACGTGAA 500

     167 nLysLysAspLysGlnLysArgThrAlaLeuHisLeuAlaSerAlaAsnG 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     501 CAAGAAGGACAAGCAAAAGAGGACTGCTCTACATCTGGCCTCTGCCAATG 550

     184 lyAsnSerGluValValLysLeuLeuLeuAspArgArgCysGlnLeuAsn 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     551 GGAATTCAGAAGTAGTAAAACTCCTGCTGGACAGACGATGTCAACTTAAT 600

     201 ValLeuAspAsnLysLysArgThrAlaLeuIleLysAlaValGlnCysGl 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     601 GTCCTTGACAACAAAAAGAGGACAGCTCTGATAAAGGCCGTACAATGCCA 650

     217 nGluAspGluCysAlaLeuMetLeuLeuGluHisGlyThrAspProAsnI 234
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     651 GGAAGATGAATGTGCGTTAATGTTGCTGGAACATGGCACTGATCCAAATA 700

     234 leProAspGluTyrGlyAsnThrThrLeuHisTyrAlaIleTyrAsnGlu 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     701 TTCCAGATGAGTATGGAAATACCACTCTGCACTACGCTATCTATAATGAA 750

     251 AspLysLeuMetAlaLysAlaLeuLeuLeuTyrGlyAlaAspIleGluSe 267
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     751 GATAAATTAATGGCCAAAGCACTGCTCTTATATGGTGCTGATATCGAATC 800

     267 rLysAsnLysHisGlyLeuThrProLeuLeuLeuGlyValHisGluGlnL 284
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     801 AAAAAACAAGCATGGCCTCACACCACTGTTACTTGGTGTACATGAGCAAA 850

     284 ysGlnGlnValValLysPheLeuIleLysLysLysAlaAsnLeuAsnAla 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     851 AACAGCAAGTCGTGAAATTTTTAATCAAGAAAAAAGCGAATTTAAATGCA 900

     301 LeuAspArgTyrGlyArgThrAlaLeuIleLeuAlaValCysCysGlySe 317
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     901 CTGGATAGATATGGAAGGACTGCTCTCATACTTGCTGTATGTTGTGGATC 950

     317 rAlaSerIleValSerLeuLeuLeuGluGlnAsnIleAspValSerSerG 334
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     951 AGCAAGTATAGTCAGCCTTCTACTTGAGCAAAATATTGATGTATCTTCTC 1000

     334 lnAspLeuSerGlyGlnThrAlaArgGluTyrAlaValSerSerHisHis 350
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1001 AAGATCTATCTGGACAGACGGCCAGAGAGTATGCTGTTTCTAGTCATCAT 1050

     351 HisValIleCysGlnLeuLeuSerAspTyrLysGluLysGlnMetLeuLy 367
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1051 CATGTAATTTGCCAGTTACTTTCTGACTACAAAGAAAAACAGATGCTAAA 1100

     367 sIleSerSerGluAsnSerAsnProGluAsnValSerArgThrArgAsnL 384
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1101 AATCTCTTCTGAAAACAGCAATCCAGAAAATGTCTCAAGAACCAGAAATA 1150
```

```
     384 ys 384
         ||
    1151 AA 1152

seq_name: gb_pat:AX316986

seq_documentation_block:
LOCUS       AX316986               1590 bp    DNA     linear   PAT 14-DEC-2001
DEFINITION  Sequence 323 from Patent WO0190152.
ACCESSION   AX316986
VERSION     AX316986.1  GI:17900055
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Frudakis,T.N., Reed,S.G., Smith,J.M., Misher,L.E., Dillon,D.C.,
            Retter,M.W., Wang,A., Skeiky,Y.A., Harlocker,S.L. and Day,C.H.
  TITLE     Compositions and methods for the therapy and diagnosis of breast
            cancer
  JOURNAL   Patent: WO 0190152-A 323 29-NOV-2001;
            CORIXA CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1. .1590
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      424 a    403 c    433 g    329 t      1 others
ORIGIN


alignment_scores:
           Quality: 2064.00             Length:    384
             Ratio:  5.375                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-810-936-304 x AX316986   ..

 Align seg 1/1  to: AX316986  from: 1  to: 1590

       1 MetValValGluValAspSerMetProAlaAlaSerSerValLysLysPr 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     436 ATGGTGGTTGAGGTTGATTCCATGCCGGCTGCTTCTTCTGTGAAGAAGCC 485

      17 oPheGlyLeuArgSerLysMetGlyLysTrpCysCysArgCysPheProC 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     486 ATTTGGTCTCAGGAGCAAGATGGGCAAGTGGTGCTGCCGTTGCTTCCCCT 535

      34 ysCysArgGluSerGlyLysSerAsnValGlyThrSerGlyAspHisAsp 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     536 GCTGCAGGGAGAGCGGCAAGAGCAACGTGGGCACTTCTGGAGACCACGAC 585

      51 AspSerAlaMetLysThrLeuArgSerLysMetGlyLysTrpCysArgHi 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     586 GACTCTGCTATGAAGACACTCAGGAGCAAGATGGGCAAGTGGTGCCGCCA 635

      67 sCysPheProCysCysArgGlySerGlyLysSerAsnValGlyAlaSerG 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     636 CTGCTTCCCCTGCTGCAGGGGGAGTGGCAAGAGCAACGTGGGCGCTTCTG 685

      84 lyAspHisAspAspSerAlaMetLysThrLeuArgAsnLysMetGlyLys 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     686 GAGACCACGACGACTCTGCTATGAAGACACTCAGGAACAAGATGGGCAAG 735

     101 TrpCysCysHisCysPheProCysCysArgGlySerGlyLysSerLysVa 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     736 TGGTGCTGCCACTGCTTCCCCTGCTGCAGGGGGAGCGGCAAGAGCAAGGT 785

     117 lGlyAlaTrpGlyAspTyrAspAspSerAlaPheMetGluProArgTyrH 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     786 GGGCGCTTGGGGAGACTACGATGACAGYGCCTTCATGGAGCCCAGGTACC 835

     134 isValArgGlyGluAspLeuAspLysLeuHisArgAlaAlaTrpTrpGly 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     836 ACGTCCGTGGAGAAGATCTGGACAAGCTCCACAGAGCTGCCTGGTGGGGT 885

     151 LysValProArgLysAspLeuIleValMetLeuArgAspThrAspValAs 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     886 AAAGTCCCCAGAAAGGATCTCATCGTCATGCTCAGGGACACTGACGTGAA 935

     167 nLysLysAspLysGlnLysArgThrAlaLeuHisLeuAlaSerAlaAsnG 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     936 CAAGAAGGACAAGCAAAAGAGGACTGCTCTACATCTGGCCTCTGCCAATG 985

     184 lyAsnSerGluValValLysLeuLeuLeuAspArgArgCysGlnLeuAsn 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     986 GGAATTCAGAAGTAGTAAAACTCCTGCTGGACAGACGATGTCAACTTAAT 1035

     201 ValLeuAspAsnLysLysArgThrAlaLeuIleLysAlaValGlnCysGl 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1036 GTCCTTGACAACAAAAAGAGGACAGCTCTGATAAAGGCCGTACAATGCCA 1085

     217 nGluAspGluCysAlaLeuMetLeuLeuGluHisGlyThrAspProAsnI 234
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1086 GGAAGATGAATGTGCGTTAATGTTGCTGGAACATGGCACTGATCCAAATA 1135

     234 leProAspGluTyrGlyAsnThrThrLeuHisTyrAlaIleTyrAsnGlu 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1136 TTCCAGATGAGTATGGAAATACCACTCTGCACTACGCTATCTATAATGAA 1185

     251 AspLysLeuMetAlaLysAlaLeuLeuLeuTyrGlyAlaAspIleGluSe 267
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1186 GATAAATTAATGGCCAAAGCACTGCTCTTATATGGTGCTGATATCGAATC 1235

     267 rLysAsnLysHisGlyLeuThrProLeuLeuLeuGlyValHisGluGlnL 284
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1236 AAAAAACAAGCATGGCCTCACACCACTGTTACTTGGTGTACATGAGCAAA 1285

     284 ysGlnGlnValValLysPheLeuIleLysLysLysAlaAsnLeuAsnAla 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1286 AACAGCAAGTCGTGAAATTTTTAATCAAGAAAAAAGCGAATTTAAATGCA 1335

     301 LeuAspArgTyrGlyArgThrAlaLeuIleLeuAlaValCysCysGlySe 317
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1336 CTGGATAGATATGGAAGGACTGCTCTCATACTTGCTGTATGTTGTGGATC 1385

     317 rAlaSerIleValSerLeuLeuLeuGluGlnAsnIleAspValSerSerG 334
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1386 AGCAAGTATAGTCAGCCTTCTACTTGAGCAAAATATTGATGTATCTTCTC 1435

     334 lnAspLeuSerGlyGlnThrAlaArgGluTyrAlaValSerSerHisHis 350
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1436 AAGATCTATCTGGACAGACGGCCAGAGAGTATGCTGTTTCTAGTCATCAT 1485

     351 HisValIleCysGlnLeuLeuSerAspTyrLysGluLysGlnMetLeuLy 367
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1486 CATGTAATTTGCCAGTTACTTTCTGACTACAAAGAAAAACAGATGCTAAA 1535

     367 sIleSerSerGluAsnSerAsnProGluAsnValSerArgThrArgAsnL 384
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1536 AATCTCTTCTGAAAACAGCAATCCAGAAAATGTCTCAAGAACCAGAAATA 1585

     384 ys 384
         ||
    1586 AA 1587

seq_name: gb_pat:AX316991

seq_documentation_block:
LOCUS       AX316991               1155 bp    DNA     linear   PAT 14-DEC-2001
DEFINITION  Sequence 328 from Patent WO0190152.
ACCESSION   AX316991
VERSION     AX316991.1  GI:17900058
KEYWORDS    .
SOURCE      human.
```

```
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Frudakis,T.N., Reed,S.G., Smith,J.M., Misher,L.E., Dillon,D.C.,
            Retter,M.W., Wang,A., Skeiky,Y.A., Harlocker,S.L. and Day,C.H.
  TITLE     Compositions and methods for the therapy and diagnosis of breast
            cancer
  JOURNAL   Patent: WO 0190152-A 328 29-NOV-2001;
            CORIXA CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1. .1155
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      346 a    253 c    296 g    260 t
ORIGIN


alignment_scores:
            Quality: 2054.00             Length:    384
              Ratio:  5.349                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 99.479

alignment_block:
 US-09-810-936-304 x AX316991   ..

 Align seg 1/1  to: AX316991  from: 1  to: 1155

       1 MetValValGluValAspSerMetProAlaAlaSerSerValLysLysPr 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 ATGGTGGTTGAGGTTGATTCCATGCCGGCTGCCTCTTCTGTGAAGAAGCC 50

      17 oPheGlyLeuArgSerLysMetGlyLysTrpCysCysArgCysPheProC 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 ATTTGGTCTCAGGAGCAAGATGGGCAAGTGGTGCTGCCGTTGCTTCCCCT 100

      34 ysCysArgGluSerGlyLysSerAsnValGlyThrSerGlyAspHisAsp 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 GCTGCAGGGAGAGCGGCAAGAGCAACGTGGGCACTTCTGGAGACCACGAC 150

      51 AspSerAlaMetLysThrLeuArgSerLysMetGlyLysTrpCysArgHi 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     151 GACTCTGCTATGAAGACACTCAGGAGCAAGATGGGCAAGTGGTGCCGCCA 200

      67 sCysPheProCysCysArgGlySerGlyLysSerAsnValGlyAlaSerG 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     201 CTGCTTCCCCTGCTGCAGGGGGAGTGGCAAGAGCAACGTGGGCGCTTCTG 250

      84 lyAspHisAspAspSerAlaMetLysThrLeuArgAsnLysMetGlyLys 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     251 GAGACCACGACGACTCTGCTATGAAGACACTCAGGAACAAGATGGGCAAG 300

     101 TrpCysCysHisCysPheProCysCysArgGlySerGlyLysSerLysVa 117
         ||||||||||||||||||||||||||||||||||:::||||||||||||
     301 TGGTGCTGCCACTGCTTCCCCTGCTGCAGGGGGAGCAGCAAGAGCAAGGT 350

     117 lGlyAlaTrpGlyAspTyrAspAspSerAlaPheMetGluProArgTyrH 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     351 GGGCGCTTGGGGAGACTACGATGACAGTGCCTTCATGGAGCCCAGGTACC 400

     134 isValArgGlyGluAspLeuAspLysLeuHisArgAlaAlaTrpTrpGly 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     401 ACGTCCGTGGAGAAGATCTGGACAAGCTCCACAGAGCTGCCTGGTGGGGT 450

     151 LysValProArgLysAspLeuIleValMetLeuArgAspThrAspValAs 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     451 AAAGTCCCCAGAAAGGATCTCATCGTCATGCTCAGGGACACTGACGTGAA 500

     167 nLysLysAspLysGlnLysArgThrAlaLeuHisLeuAlaSerAlaAsnG 184
         ||||:::||||||||||||||||||||||||||||||||||||||||||
     501 CAAGCAGGACAAGCAAAAGAGGACTGCTCTACATCTGGCCTCTGCCAATG 550

     184 lyAsnSerGluValValLysLeuLeuLeuAspArgArgCysGlnLeuAsn 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     551 GGAATTCAGAAGTAGTAAAACTCCTGCTGGACAGACGATGTCAACTTAAT 600

     201 ValLeuAspAsnLysLysArgThrAlaLeuIleLysAlaValGlnCysGl 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     601 GTCCTTGACAACAAAAAGAGGACAGCTCTGATAAAGGCCGTACAATGCCA 650

     217 nGluAspGluCysAlaLeuMetLeuLeuGluHisGlyThrAspProAsnI 234
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     651 GGAAGATGAATGTGCGTTAATGTTGCTGGAACATGGCACTGATCCAAATA 700

     234 leProAspGluTyrGlyAsnThrThrLeuHisTyrAlaIleTyrAsnGlu 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     701 TTCCAGATGAGTATGGAAATACCACTCTGCACTACGCTATCTATAATGAA 750

     251 AspLysLeuMetAlaLysAlaLeuLeuLeuTyrGlyAlaAspIleGluSe 267
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     751 GATAAATTAATGGCCAAAGCACTGCTCTTATATGGTGCTGATATCGAATC 800

     267 rLysAsnLysHisGlyLeuThrProLeuLeuLeuGlyValHisGluGlnL 284
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     801 AAAAAACAAGCATGGCCTCACACCACTGTTACTTGGTGTACATGAGCAAA 850

     284 ysGlnGlnValValLysPheLeuIleLysLysLysAlaAsnLeuAsnAla 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     851 AACAGCAAGTCGTGAAATTTTTAATTAAGAAAAAAGCGAATTTAAATGCA 900

     301 LeuAspArgTyrGlyArgThrAlaLeuIleLeuAlaValCysCysGlySe 317
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     901 CTGGATAGATATGGAAGGACTGCTCTCATACTTGCTGTATGTTGTGGATC 950

     317 rAlaSerIleValSerLeuLeuLeuGluGlnAsnIleAspValSerSerG 334
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     951 AGCAAGTATAGTCAGCCTTCTACTTGAGCAAAATATTGATGTATCTTCTC 1000

     334 lnAspLeuSerGlyGlnThrAlaArgGluTyrAlaValSerSerHisHis 350
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1001 AAGATCTATCTGGACAGACGGCCAGAGAGTATGCTGTTTCTAGTCATCAT 1050

     351 HisValIleCysGlnLeuLeuSerAspTyrLysGluLysGlnMetLeuLy 367
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1051 CATGTAATTTGCCAGTTACTTTCTGACTACAAAGAAAAACAGATGCTAAA 1100

     367 sIleSerSerGluAsnSerAsnProGluAsnValSerArgThrArgAsnL 384
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1101 AATCTCTTCTGAAAACAGCAATCCAGAAAATGTCTCAAGAACCAGAAATA 1150

      384 ys 384
          ||
    1151 AA 1152

seq_name: gb_pat:AX106593

seq_documentation_block:
LOCUS       AX106593               2000 bp    DNA     linear   PAT 30-APR-2001
DEFINITION  Sequence 374 from Patent WO0125272.
ACCESSION   AX106593
VERSION     AX106593.1  GI:13922264
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2000)
  AUTHORS   Xu,J., Skeiky,Y.A., Reed,S.G. and Cheever,M.A.
  TITLE     Compositions and methods for therapy and diagnosis of prostate
            cancer
  JOURNAL   Patent: WO 0125272-A 374 12-APR-2001;
            CORIXA CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1. .2000
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
```

```
BASE COUNT      698 a    388 c    489 g    425 t
ORIGIN


alignment_scores:
            Quality: 2024.00             Length:    376
              Ratio:  5.383               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-810-936-304 x AX106593   ..

 Align seg 1/1  to: AX106593  from: 1  to: 2000

       1 MetValValGluValAspSerMetProAlaAlaSerSerValLysLysPr 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 ATGGTGGTTGAGGTTGATTCCATGCCGGCTGCCTCTTCTGTGAAGAAGCC 50

      17 oPheGlyLeuArgSerLysMetGlyLysTrpCysCysArgCysPheProC 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 ATTTGGTCTCAGGAGCAAGATGGGCAAGTGGTGCTGCCGTTGCTTCCCCT 100

      34 ysCysArgGluSerGlyLysSerAsnValGlyThrSerGlyAspHisAsp 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 GCTGCAGGGAGAGCGGCAAGAGCAACGTGGGCACTTCTGGAGACCACGAC 150

      51 AspSerAlaMetLysThrLeuArgSerLysMetGlyLysTrpCysArgHi 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     151 GACTCTGCTATGAAGACACTCAGGAGCAAGATGGGCAAGTGGTGCCGCCA 200

      67 sCysPheProCysCysArgGlySerGlyLysSerAsnValGlyAlaSerG 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     201 CTGCTTCCCCTGCTGCAGGGGGAGTGGCAAGAGCAACGTGGGCGCTTCTG 250

      84 lyAspHisAspAspSerAlaMetLysThrLeuArgAsnLysMetGlyLys 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     251 GAGACCACGACGACTCTGCTATGAAGACACTCAGGAACAAGATGGGCAAG 300

     101 TrpCysCysHisCysPheProCysCysArgGlySerGlyLysSerLysVa 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     301 TGGTGCTGCCACTGCTTCCCCTGCTGCAGGGGGAGCGGCAAGAGCAAGGT 350

     117 lGlyAlaTrpGlyAspTyrAspAspSerAlaPheMetGluProArgTyrH 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     351 GGGCGCTTGGGGAGACTACGATGACAGTGCCTTCATGGAGCCCAGGTACC 400

     134 isValArgGlyGluAspLeuAspLysLeuHisArgAlaAlaTrpTrpGly 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     401 ACGTCCGTGGAGAAGATCTGGACAAGCTCCACAGAGCTGCCTGGTGGGGT 450

     151 LysValProArgLysAspLeuIleValMetLeuArgAspThrAspValAs 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     451 AAAGTCCCCAGAAAGGATCTCATCGTCATGCTCAGGGACACTGACGTGAA 500

     167 nLysLysAspLysGlnLysArgThrAlaLeuHisLeuAlaSerAlaAsnG 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     501 CAAGAAGGACAAGCAAAAGAGGACTGCTCTACATCTGGCCTCTGCCAATG 550

     184 lyAsnSerGluValValLysLeuLeuLeuAspArgArgCysGlnLeuAsn 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     551 GGAATTCAGAAGTAGTAAAACTCCTGCTGGACAGACGATGTCAACTTAAT 600

     201 ValLeuAspAsnLysLysArgThrAlaLeuIleLysAlaValGlnCysGl 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     601 GTCCTTGACAACAAAAAGAGGACAGCTCTGATAAAGGCCGTACAATGCCA 650

     217 nGluAspGluCysAlaLeuMetLeuLeuGluHisGlyThrAspProAsnI 234
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     651 GGAAGATGAATGTGCGTTAATGTTGCTGGAACATGGCACTGATCCAAATA 700

     234 leProAspGluTyrGlyAsnThrThrLeuHisTyrAlaIleTyrAsnGlu 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     701 TTCCAGATGAGTATGGAAATACCACTCTGCACTACGCTATCTATAATGAA 750

     251 AspLysLeuMetAlaLysAlaLeuLeuLeuTyrGlyAlaAspIleGluSe 267
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     751 GATAAATTAATGGCCAAAGCACTGCTCTTATATGGTGCTGATATCGAATC 800

     267 rLysAsnLysHisGlyLeuThrProLeuLeuLeuGlyValHisGluGlnL 284
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     801 AAAAAACAAGCATGGCCTCACACCACTGTTACTTGGTGTACATGAGCAAA 850

     284 ysGlnGlnValValLysPheLeuIleLysLysLysAlaAsnLeuAsnAla 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     851 AACAGCAAGTCGTGAAATTTTTAATCAAGAAAAAAGCGAATTTAAATGCA 900

     301 LeuAspArgTyrGlyArgThrAlaLeuIleLeuAlaValCysCysGlySe 317
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     901 CTGGATAGATATGGAAGGACTGCTCTCATACTTGCTGTATGTTGTGGATC 950

     317 rAlaSerIleValSerLeuLeuLeuGluGlnAsnIleAspValSerSerG 334
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     951 AGCAAGTATAGTCAGCCTTCTACTTGAGCAAAATATTGATGTATCTTCTC 1000

     334 lnAspLeuSerGlyGlnThrAlaArgGluTyrAlaValSerSerHisHis 350
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1001 AAGATCTATCTGGACAGACGGCCAGAGAGTATGCTGTTTCTAGTCATCAT 1050

     351 HisValIleCysGlnLeuLeuSerAspTyrLysGluLysGlnMetLeuLy 367
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1051 CATGTAATTTGCCAGTTACTTTCTGACTACAAAGAAAAACAGATGCTAAA 1100

     367 sIleSerSerGluAsnSerAsnProGlu 376
         |||||||||||||||||||||||||||
    1101 AATCTCTTCTGAAAACAGCAATCCAGAA 1128

seq_name: gb_pat:AX140884

seq_documentation_block:
LOCUS       AX140884               2000 bp    DNA     linear   PAT 31-MAY-2001
DEFINITION  Sequence 374 from Patent WO0134802.
ACCESSION   AX140884
VERSION     AX140884.1  GI:14280987
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2000)
  AUTHORS   Xu,J., Dillon,D.C., Mitcham,J.L., Harlocker,S.L., Jiang,Y.,
            Reed,S.G., Kalos,M.D., Retter,M.W., Stolk,J.A., Day,C.H.,
            Skeiky,Y.A. and Wang,A.
  TITLE     Compositions and methods for the therapy and diagnosis of prostate
            cancer
  JOURNAL   Patent: WO 0134802-A 374 17-MAY-2001;
            CORIXA CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1. .2000
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      698 a    388 c    489 g    425 t
ORIGIN


alignment_scores:
            Quality: 2024.00             Length:    376
              Ratio:  5.383               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-810-936-304 x AX140884   ..

 Align seg 1/1  to: AX140884  from: 1  to: 2000

       1 MetValValGluValAspSerMetProAlaAlaSerSerValLysLysPr 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
  1 ATGGTGGTTGAGGTTGATTCCATGCCGGCTGCCTCTTCTGTGAAGAAGCC 50

 17 oPheGlyLeuArgSerLysMetGlyLysTrpCysCysArgCysPheProC 34
    ||||||||||||||||||||||||||||||||||||||||||||||||||
 51 ATTTGGTCTCAGGAGCAAGATGGGCAAGTGGTGCTGCCGTTGCTTCCCCT 100

 34 ysCysArgGluSerGlyLysSerAsnValGlyThrSerGlyAspHisAsp 50
    ||||||||||||||||||||||||||||||||||||||||||||||||||
101 GCTGCAGGGAGAGCGGCAAGAGCAACGTGGGCACTTCTGGAGACCACGAC 150

 51 AspSerAlaMetLysThrLeuArgSerLysMetGlyLysTrpCysArgHi 67
    ||||||||||||||||||||||||||||||||||||||||||||||||||
151 GACTCTGCTATGAAGACACTCAGGAGCAAGATGGGCAAGTGGTGCCGCCA 200

 67 sCysPheProCysCysArgGlySerGlyLysSerAsnValGlyAlaSerG 84
    ||||||||||||||||||||||||||||||||||||||||||||||||||
201 CTGCTTCCCCTGCTGCAGGGGGAGTGGCAAGAGCAACGTGGGCGCTTCTG 250

 84 lyAspHisAspAspSerAlaMetLysThrLeuArgAsnLysMetGlyLys 100
    ||||||||||||||||||||||||||||||||||||||||||||||||||
251 GAGACCACGACGACTCTGCTATGAAGACACTCAGGAACAAGATGGGCAAG 300

101 TrpCysCysHisCysPheProCysCysArgGlySerGlyLysSerLysVa 117
    ||||||||||||||||||||||||||||||||||||||||||||||||||
301 TGGTGCTGCCACTGCTTCCCCTGCTGCAGGGGGAGCGGCAAGAGCAAGGT 350

117 lGlyAlaTrpGlyAspTyrAspAspSerAlaPheMetGluProArgTyrH 134
    ||||||||||||||||||||||||||||||||||||||||||||||||||
351 GGGCGCTTGGGGAGACTACGATGACAGTGCCTTCATGGAGCCCAGGTACC 400

134 isValArgGlyGluAspLeuAspLysLeuHisArgAlaAlaTrpTrpGly 150
    ||||||||||||||||||||||||||||||||||||||||||||||||||
401 ACGTCCGTGGAGAAGATCTGGACAAGCTCCACAGAGCTGCCTGGTGGGGT 450

151 LysValProArgLysAspLeuIleValMetLeuArgAspThrAspValAs 167
    ||||||||||||||||||||||||||||||||||||||||||||||||||
451 AAAGTCCCCAGAAAGGATCTCATCGTCATGCTCAGGGACACTGACGTGAA 500

167 nLysLysAspLysGlnLysArgThrAlaLeuHisLeuAlaSerAlaAsnG 184
    ||||||||||||||||||||||||||||||||||||||||||||||||||
501 CAAGAAGGACAAGCAAAAGAGGACTGCTCTACATCTGGCCTCTGCCAATG 550

184 lyAsnSerGluValValLysLeuLeuLeuAspArgArgCysGlnLeuAsn 200
    ||||||||||||||||||||||||||||||||||||||||||||||||||
551 GGAATTCAGAAGTAGTAAAACTCCTGCTGGACAGACGATGTCAACTTAAT 600

201 ValLeuAspAsnLysLysArgThrAlaLeuIleLysAlaValGlnCysGl 217
    ||||||||||||||||||||||||||||||||||||||||||||||||||
601 GTCCTTGACAACAAAAAGAGGACAGCTCTGATAAAGGCCGTACAATGCCA 650

217 nGluAspGluCysAlaLeuMetLeuLeuGluHisGlyThrAspProAsnI 234
    ||||||||||||||||||||||||||||||||||||||||||||||||||
651 GGAAGATGAATGTGCGTTAATGTTGCTGGAACATGGCACTGATCCAAATA 700

234 leProAspGluTyrGlyAsnThrThrLeuHisTyrAlaIleTyrAsnGlu 250
    ||||||||||||||||||||||||||||||||||||||||||||||||||
701 TTCCAGATGAGTATGGAAATACCACTCTGCACTACGCTATCTATAATGAA 750

251 AspLysLeuMetAlaLysAlaLeuLeuLeuTyrGlyAlaAspIleGluSe 267
    ||||||||||||||||||||||||||||||||||||||||||||||||||
751 GATAAATTAATGGCCAAAGCACTGCTCTTATATGGTGCTGATATCGAATC 800

267 rLysAsnLysHisGlyLeuThrProLeuLeuLeuGlyValHisGluGlnL 284
    ||||||||||||||||||||||||||||||||||||||||||||||||||
801 AAAAAACAAGCATGGCCTCACACCACTGTTACTTGGTGTACATGAGCAAA 850

284 ysGlnGlnValValLysPheLeuIleLysLysLysAlaAsnLeuAsnAla 300
    ||||||||||||||||||||||||||||||||||||||||||||||||||
851 AACAGCAAGTCGTGAAATTTTTAATCAAGAAAAAAGCGAATTTAAATGCA 900

301 LeuAspArgTyrGlyArgThrAlaLeuIleLeuAlaValCysCysGlySe 317
    ||||||||||||||||||||||||||||||||||||||||||||||||||
901 CTGGATAGATATGGAAGGACTGCTCTCATACTTGCTGTATGTTGTGGATC 950

317 rAlaSerIleValSerLeuLeuLeuGluGlnAsnIleAspValSerSerG 334
    ||||||||||||||||||||||||||||||||||||||||||||||||||
951 AGCAAGTATAGTCAGCCTTCTACTTGAGCAAAATATTGATGTATCTTCTC 1000

334 lnAspLeuSerGlyGlnThrAlaArgGluTyrAlaValSerSerHisHis 350
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1001 AAGATCTATCTGGACAGACGGCCAGAGAGTATGCTGTTTCTAGTCATCAT 1050

351 HisValIleCysGlnLeuLeuSerAspTyrLysGluLysGlnMetLeuLy 367
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1051 CATGTAATTTGCCAGTTACTTTCTGACTACAAAGAAAAACAGATGCTAAA 1100

367 sIleSerSerGluAsnSerAsnProGlu 376
    |||||||||||||||||||||||||||
1101 AATCTCTTCTGAAAACAGCAATCCAGAA 1128
```

```
seq_name: gb_pat:AX200744

seq_documentation_block:
LOCUS       AX200744               2000 bp    DNA     linear   PAT 29-AUG-2001
DEFINITION  Sequence 374 from Patent WO0151633.
ACCESSION   AX200744
VERSION     AX200744.1  GI:15390637
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2000)
  AUTHORS   Xu,J., Dillon,D.C., Mitcham,J.L., Harlocker,S.L., Jiang,Y.,
            Reed,S.G., Kalos,M.D., Fanger,G.R., Day,C.H., Retter,M.W.,
            Stolk,J.A., Skeiky,Y.A., Wang,A. and Meagher,M.J.
  TITLE     Compositions and methods for the therapy and diagnosis of prostate
            cancer
  JOURNAL   Patent: WO 0151633-A 374 19-JUL-2001;
            CORIXA CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1. .2000
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      698 a    388 c    489 g    425 t
ORIGIN


alignment_scores:
            Quality: 2024.00             Length:    376
              Ratio:  5.383                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-810-936-304 x AX200744   ..

 Align seg 1/1  to: AX200744  from: 1  to: 2000

       1 MetValValGluValAspSerMetProAlaAlaSerSerValLysLysPr 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 ATGGTGGTTGAGGTTGATTCCATGCCGGCTGCCTCTTCTGTGAAGAAGCC 50

      17 oPheGlyLeuArgSerLysMetGlyLysTrpCysCysArgCysPheProC 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 ATTTGGTCTCAGGAGCAAGATGGGCAAGTGGTGCTGCCGTTGCTTCCCCT 100

      34 ysCysArgGluSerGlyLysSerAsnValGlyThrSerGlyAspHisAsp 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 GCTGCAGGGAGAGCGGCAAGAGCAACGTGGGCACTTCTGGAGACCACGAC 150

      51 AspSerAlaMetLysThrLeuArgSerLysMetGlyLysTrpCysArgHi 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     151 GACTCTGCTATGAAGACACTCAGGAGCAAGATGGGCAAGTGGTGCCGCCA 200

      67 sCysPheProCysCysArgGlySerGlyLysSerAsnValGlyAlaSerG 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
201 CTGCTTCCCCTGCTGCAGGGGGAGTGGCAAGAGCAACGTGGGCGCTTCTG 250

 84 lyAspHisAspAspSerAlaMetLysThrLeuArgAsnLysMetGlyLys 100
    !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
251 GAGACCACGACGACTCTGCTATGAAGACACTCAGGAACAAGATGGGCAAG 300

101 TrpCysCysHisCysPheProCysCysArgGlySerGlyLysSerLysVa 117
    !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
301 TGGTGCTGCCACTGCTTCCCCTGCTGCAGGGGGAGCGGCAAGAGCAAGGT 350

117 lGlyAlaTrpGlyAspTyrAspAspSerAlaPheMetGluProArgTyrH 134
    !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
351 GGGCGCTTGGGGAGACTACGATGACAGTGCCTTCATGGAGCCCAGGTACC 400

134 isValArgGlyGluAspLeuAspLysLeuHisArgAlaAlaTrpTrpGly 150
    !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
401 ACGTCCGTGGAGAAGATCTGGACAAGCTCCACAGAGCTGCCTGGTGGGGT 450

151 LysValProArgLysAspLeuIleValMetLeuArgAspThrAspValAs 167
    !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
451 AAAGTCCCCAGAAAGGATCTCATCGTCATGCTCAGGGACACTGACGTGAA 500

167 nLysLysAspLysGlnLysArgThrAlaLeuHisLeuAlaSerAlaAsnG 184
    !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
501 CAAGAAGGACAAGCAAAAGAGGACTGCTCTACATCTGGCCTCTGCCAATG 550

184 lyAsnSerGluValValLysLeuLeuLeuAspArgArgCysGlnLeuAsn 200
    !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
551 GGAATTCAGAAGTAGTAAAACTCCTGCTGGACAGACGATGTCAACTTAAT 600

201 ValLeuAspAsnLysLysArgThrAlaLeuIleLysAlaValGlnCysGl 217
    !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
601 GTCCTTGACAACAAAAAGAGGACAGCTCTGATAAAGGCCGTACAATGCCA 650

217 nGluAspGluCysAlaLeuMetLeuLeuGluHisGlyThrAspProAsnI 234
    !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
651 GGAAGATGAATGTGCGTTAATGTTGCTGGAACATGGCACTGATCCAAATA 700

234 leProAspGluTyrGlyAsnThrThrLeuHisTyrAlaIleTyrAsnGlu 250
    !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
701 TTCCAGATGAGTATGGAAATACCACTCTGCACTACGCTATCTATAATGAA 750

251 AspLysLeuMetAlaLysAlaLeuLeuLeuTyrGlyAlaAspIleGluSe 267
    !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
751 GATAAATTAATGGCCAAAGCACTGCTCTTATATGGTGCTGATATCGAATC 800

267 rLysAsnLysHisGlyLeuThrProLeuLeuLeuGlyValHisGluGlnL 284
    !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
801 AAAAAACAAGCATGGCCTCACACCACTGTTACTTGGTGTACATGAGCAAA 850

284 ysGlnGlnValValLysPheLeuIleLysLysLysAlaAsnLeuAsnAla 300
    !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
851 AACAGCAAGTCGTGAAATTTTTAATCAAGAAAAAAGCGAATTTAAATGCA 900

301 LeuAspArgTyrGlyArgThrAlaLeuIleLeuAlaValCysCysGlySe 317
    !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
901 CTGGATAGATATGGAAGGACTGCTCTCATACTTGCTGTATGTTGTGGATC 950

317 rAlaSerIleValSerLeuLeuLeuGluGlnAsnIleAspValSerSerG 334
    !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
951 AGCAAGTATAGTCAGCCTTCTACTTGAGCAAAATATTGATGTATCTTCTC 1000

334 lnAspLeuSerGlyGlnThrAlaArgGluTyrAlaValSerSerHisHis 350
    !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
1001 AAGATCTATCTGGACAGACGGCCAGAGAGTATGCTGTTTCTAGTCATCAT 1050

351 HisValIleCysGlnLeuLeuSerAspTyrLysGluLysGlnMetLeuLy 367
    !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
1051 CATGTAATTTGCCAGTTACTTTCTGACTACAAAGAAAAACAGATGCTAAA 1100

367 sIleSerSerGluAsnSerAsnProGlu 376
    !!!!!!!!!!!!!!!!!!!!!!!!!!!!
1101 AATCTCTTCTGAAAACAGCAATCCAGAA 1128
```

```
seq_name: gb_pat:AX267400

seq_documentation_block:
LOCUS       AX267400               2000 bp    DNA     linear   PAT 26-OCT-2001
DEFINITION  Sequence 374 from Patent WO0173032.
ACCESSION   AX267400
VERSION     AX267400.1  GI:16516170
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Xu,J., Dillon,D.C., Mitcham,J.L., Harlocker,S.L., Jiang,Y.,
            Kalos,M.D., Fanger,G.R., Retter,M.W., Stolk,J.A., Day,C.H.,
            Vedvick,T.S., Carter,D., Li,S.X., Wang,A., Skeiky,Y.A., Hepler,W.T.
            and Henderson,R.A.
  TITLE     Compositions and methods for the therapy and diagnosis of prostate
            cancer
  JOURNAL   Patent: WO 0173032-A 374 04-OCT-2001;
            CORIXA CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1. .2000
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      698 a    388 c    489 g    425 t
ORIGIN


alignment_scores:
            Quality: 2024.00             Length:    376
              Ratio:  5.383                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-810-936-304 x AX267400   ..

 Align seg 1/1  to: AX267400  from: 1  to: 2000

       1 MetValValGluValAspSerMetProAlaAlaSerSerValLysLysPr 17
         !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
       1 ATGGTGGTTGAGGTTGATTCCATGCCGGCTGCCTCTTCTGTGAAGAAGCC 50

      17 oPheGlyLeuArgSerLysMetGlyLysTrpCysCysArgCysPheProC 34
         !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
      51 ATTTGGTCTCAGGAGCAAGATGGGCAAGTGGTGCTGCCGTTGCTTCCCCT 100

      34 ysCysArgGluSerGlyLysSerAsnValGlyThrSerGlyAspHisAsp 50
         !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
     101 GCTGCAGGGAGAGCGGCAAGAGCAACGTGGGCACTTCTGGAGACCACGAC 150

      51 AspSerAlaMetLysThrLeuArgSerLysMetGlyLysTrpCysArgHi 67
         !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
     151 GACTCTGCTATGAAGACACTCAGGAGCAAGATGGGCAAGTGGTGCCGCCA 200

      67 sCysPheProCysCysArgGlySerGlyLysSerAsnValGlyAlaSerG 84
         !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
     201 CTGCTTCCCCTGCTGCAGGGGGAGTGGCAAGAGCAACGTGGGCGCTTCTG 250

      84 lyAspHisAspAspSerAlaMetLysThrLeuArgAsnLysMetGlyLys 100
         !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
     251 GAGACCACGACGACTCTGCTATGAAGACACTCAGGAACAAGATGGGCAAG 300

     101 TrpCysCysHisCysPheProCysCysArgGlySerGlyLysSerLysVa 117
         !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
     301 TGGTGCTGCCACTGCTTCCCCTGCTGCAGGGGGAGCGGCAAGAGCAAGGT 350

     117 lGlyAlaTrpGlyAspTyrAspAspSerAlaPheMetGluProArgTyrH 134
         !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
     351 GGGCGCTTGGGGAGACTACGATGACAGTGCCTTCATGGAGCCCAGGTACC 400

     134 isValArgGlyGluAspLeuAspLysLeuHisArgAlaAlaTrpTrpGly 150
```

```
     ||||||||||||||||||||||||||||||||||||||||||||||||||
401  ACGTCCGTGGAGAAGATCTGGACAAGCTCCACAGAGCTGCCTGGTGGGGT 450

151  LysValProArgLysAspLeuIleValMetLeuArgAspThrAspValAs 167
     ||||||||||||||||||||||||||||||||||||||||||||||||||
451  AAAGTCCCCAGAAAGGATCTCATCGTCATGCTCAGGGACACTGACGTGAA 500

167  nLysLysAspLysGlnLysArgThrAlaLeuHisLeuAlaSerAlaAsnG 184
     ||||||||||||||||||||||||||||||||||||||||||||||||||
501  CAAGAAGGACAAGCAAAAGAGGACTGCTCTACATCTGGCCTCTGCCAATG 550

184  lyAsnSerGluValValLysLeuLeuLeuAspArgArgCysGlnLeuAsn 200
     ||||||||||||||||||||||||||||||||||||||||||||||||||
551  GGAATTCAGAAGTAGTAAAACTCCTGCTGGACAGACGATGTCAACTTAAT 600

201  ValLeuAspAsnLysLysArgThrAlaLeuIleLysAlaValGlnCysGl 217
     ||||||||||||||||||||||||||||||||||||||||||||||||||
601  GTCCTTGACAACAAAAAGAGGACAGCTCTGATAAAGGCCGTACAATGCCA 650

217  nGluAspGluCysAlaLeuMetLeuLeuGluHisGlyThrAspProAsnI 234
     ||||||||||||||||||||||||||||||||||||||||||||||||||
651  GGAAGATGAATGTGCGTTAATGTTGCTGGAACATGGCACTGATCCAAATA 700

234  leProAspGluTyrGlyAsnThrThrLeuHisTyrAlaIleTyrAsnGlu 250
     ||||||||||||||||||||||||||||||||||||||||||||||||||
701  TTCCAGATGAGTATGGAAATACCACTCTGCACTACGCTATCTATAATGAA 750

251  AspLysLeuMetAlaLysAlaLeuLeuLeuTyrGlyAlaAspIleGluSe 267
     ||||||||||||||||||||||||||||||||||||||||||||||||||
751  GATAAATTAATGGCCAAAGCACTGCTCTTATATGGTGCTGATATCGAATC 800

267  rLysAsnLysHisGlyLeuThrProLeuLeuLeuGlyValHisGluGlnL 284
     ||||||||||||||||||||||||||||||||||||||||||||||||||
801  AAAAAACAAGCATGGCCTCACACCACTGTTACTTGGTGTACATGAGCAAA 850

284  ysGlnGlnValValLysPheLeuIleLysLysLysAlaAsnLeuAsnAla 300
     ||||||||||||||||||||||||||||||||||||||||||||||||||
851  AACAGCAAGTCGTGAAATTTTTAATCAAGAAAAAAGCGAATTTAAATGCA 900

301  LeuAspArgTyrGlyArgThrAlaLeuIleLeuAlaValCysCysGlySe 317
     ||||||||||||||||||||||||||||||||||||||||||||||||||
901  CTGGATAGATATGGAAGGACTGCTCTCATACTTGCTGTATGTTGTGGATC 950

317  rAlaSerIleValSerLeuLeuLeuGluGlnAsnIleAspValSerSerG 334
     ||||||||||||||||||||||||||||||||||||||||||||||||||
951  AGCAAGTATAGTCAGCCTTCTACTTGAGCAAAATATTGATGTATCTTCTC 1000

334  lnAspLeuSerGlyGlnThrAlaArgGluTyrAlaValSerSerHisHis 350
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1001 AAGATCTATCTGGACAGACGGCCAGAGAGTATGCTGTTTCTAGTCATCAT 1050

351  HisValIleCysGlnLeuLeuSerAspTyrLysGluLysGlnMetLeuLy 367
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1051 CATGTAATTTGCCAGTTACTTTCTGACTACAAAGAAAAACAGATGCTAAA 1100

367  sIleSerSerGluAsnSerAsnProGlu 376
     |||||||||||||||||||||||||||
1101 AATCTCTTCTGAAAACAGCAATCCAGAA 1128

seq_name: gb_pat:AX282957

seq_documentation_block:
LOCUS       AX282957               2000 bp    DNA     linear   PAT 02-NOV-2001
DEFINITION  Sequence 6 from Patent WO0175171.
ACCESSION   AX282957
VERSION     AX282957.1  GI:16609897
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Houghton,R.L., Dillon,D.C., Molesh,D.A., Xu,J., Zehentner,B. and
            Persing,D.H.
  TITLE     Methods, compositions and kits for the detection and monitoring of
            breast cancer
  JOURNAL   Patent: WO 0175171-A 6 11-OCT-2001;
            CORIXA CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1. .2000
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      698 a    388 c    489 g    425 t
ORIGIN


alignment_scores:
            Quality: 2024.00            Length:    376
              Ratio:  5.383               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-810-936-304 x AX282957   ..

 Align seg 1/1  to: AX282957  from: 1  to: 2000

       1 MetValValGluValAspSerMetProAlaAlaSerSerValLysLysPr 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 ATGGTGGTTGAGGTTGATTCCATGCCGGCTGCCTCTTCTGTGAAGAAGCC 50

      17 oPheGlyLeuArgSerLysMetGlyLysTrpCysCysArgCysPheProC 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 ATTTGGTCTCAGGAGCAAGATGGGCAAGTGGTGCTGCCGTTGCTTCCCCT 100

      34 ysCysArgGluSerGlyLysSerAsnValGlyThrSerGlyAspHisAsp 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 GCTGCAGGGAGAGCGGCAAGAGCAACGTGGGCACTTCTGGAGACCACGAC 150

      51 AspSerAlaMetLysThrLeuArgSerLysMetGlyLysTrpCysArgHi 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     151 GACTCTGCTATGAAGACACTCAGGAGCAAGATGGGCAAGTGGTGCCGCCA 200

      67 sCysPheProCysCysArgGlySerGlyLysSerAsnValGlyAlaSerG 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     201 CTGCTTCCCCTGCTGCAGGGGGAGTGGCAAGAGCAACGTGGGCGCTTCTG 250

      84 lyAspHisAspAspSerAlaMetLysThrLeuArgAsnLysMetGlyLys 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     251 GAGACCACGACGACTCTGCTATGAAGACACTCAGGAACAAGATGGGCAAG 300

     101 TrpCysCysHisCysProCysCysArgGlySerGlyLysSerLysVa 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     301 TGGTGCTGCCACTGCTTCCCCTGCTGCAGGGGGAGCGGCAAGAGCAAGGT 350

     117 lGlyAlaTrpGlyAspTyrAspAspSerAlaPheMetGluProArgTyrH 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     351 GGGCGCTTGGGGAGACTACGATGACAGTGCCTTCATGGAGCCCAGGTACC 400

     134 isValArgGlyGluAspLeuAspLysLeuHisArgAlaAlaTrpTrpGly 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     401 ACGTCCGTGGAGAAGATCTGGACAAGCTCCACAGAGCTGCCTGGTGGGGT 450

     151 LysValProArgLysAspLeuIleValMetLeuArgAspThrAspValAs 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     451 AAAGTCCCCAGAAAGGATCTCATCGTCATGCTCAGGGACACTGACGTGAA 500

     167 nLysLysAspLysGlnLysArgThrAlaLeuHisLeuAlaSerAlaAsnG 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     501 CAAGAAGGACAAGCAAAAGAGGACTGCTCTACATCTGGCCTCTGCCAATG 550

     184 lyAsnSerGluValValLysLeuLeuLeuAspArgArgCysGlnLeuAsn 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     551 GGAATTCAGAAGTAGTAAAACTCCTGCTGGACAGACGATGTCAACTTAAT 600

     201 ValLeuAspAsnLysLysArgThrAlaLeuIleLysAlaValGlnCysGl 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
     601 GTCCTTGACAACAAAAAGAGGACAGCTCTGATAAAGGCCGTACAATGCCA 650

     217 nGluAspGluCysAlaLeuMetLeuLeuGluHisGlyThrAspProAsnI 234
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     651 GGAAGATGAATGTGCGTTAATGTTGCTGGAACATGGCACTGATCCAAATA 700

     234 leProAspGluTyrGlyAsnThrThrLeuHisTyrAlaIleTyrAsnGlu 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     701 TTCCAGATGAGTATGGAAATACCACTCTGCACTACGCTATCTATAATGAA 750

     251 AspLysLeuMetAlaLysAlaLeuLeuLeuTyrGlyAlaAspIleGluSe 267
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     751 GATAAATTAATGGCCAAAGCACTGCTCTTATATGGTGCTGATATCGAATC 800

     267 rLysAsnLysHisGlyLeuThrProLeuLeuLeuGlyValHisGluGlnL 284
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     801 AAAAAACAAGCATGGCCTCACACCACTGTTACTTGGTGTACATGAGCAAA 850

     284 ysGlnGlnValValLysPheLeuIleLysLysLysAlaAsnLeuAsnAla 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     851 AACAGCAAGTCGTGAAATTTTTAATCAAGAAAAAAGCGAATTTAAATGCA 900

     301 LeuAspArgTyrGlyArgThrAlaLeuIleLeuAlaValCysCysGlySe 317
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     901 CTGGATAGATATGGAAGGACTGCTCTCATACTTGCTGTATGTTGTGGATC 950

     317 rAlaSerIleValSerLeuLeuLeuGluGlnAsnIleAspValSerSerG 334
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     951 AGCAAGTATAGTCAGCCTTCTACTTGAGCAAAATATTGATGTATCTTCTC 1000

     334 lnAspLeuSerGlyGlnThrAlaArgGluTyrAlaValSerSerHisHis 350
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1001 AAGATCTATCTGGACAGACGGCCAGAGAGTATGCTGTTTCTAGTCATCAT 1050

     351 HisValIleCysGlnLeuLeuSerAspTyrLysGluLysGlnMetLeuLy 367
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1051 CATGTAATTTGCCAGTTACTTTCTGACTACAAAGAAAAACAGATGCTAAA 1100

     367 sIleSerSerGluAsnSerAsnProGlu 376
         |||||||||||||||||||||||||||
    1101 AATCTCTTCTGAAAACAGCAATCCAGAA 1128

seq_name: gb_pat:AX316965

seq_documentation_block:
LOCUS       AX316965               2000 bp    DNA     linear   PAT 14-DEC-2001
DEFINITION  Sequence 302 from Patent WO0190152.
ACCESSION   AX316965
VERSION     AX316965.1  GI:17900044
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Frudakis,T.N., Reed,S.G., Smith,J.M., Misher,L.E., Dillon,D.C.,
            Retter,M.W., Wang,A., Skeiky,Y.A., Harlocker,S.L. and Day,C.H.
  TITLE     Compositions and methods for the therapy and diagnosis of breast
            cancer
  JOURNAL   Patent: WO 0190152-A 302 29-NOV-2001;
            CORIXA CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1. .2000
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      698 a    388 c    489 g    425 t
ORIGIN


alignment_scores:
            Quality: 2024.00            Length:    376
              Ratio:  5.383              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000
```

```
alignment_block:
 US-09-810-936-304 x AX316965   ..

 Align seg 1/1  to: AX316965  from: 1  to: 2000

       1 MetValValGluValAspSerMetProAlaAlaSerSerValLysLysPr 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 ATGGTGGTTGAGGTTGATTCCATGCCGGCTGCCTCTTCTGTGAAGAAGCC 50

      17 oPheGlyLeuArgSerLysMetGlyLysTrpCysCysArgCysPheProC 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 ATTTGGTCTCAGGAGCAAGATGGGCAAGTGGTGCTGCCGTTGCTTCCCCT 100

      34 ysCysArgGluSerGlyLysSerAsnValGlyThrSerGlyAspHisAsp 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 GCTGCAGGGAGAGCGGCAAGAGCAACGTGGGCACTTCTGGAGACCACGAC 150

      51 AspSerAlaMetLysThrLeuArgSerLysMetGlyLysTrpCysArgHi 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     151 GACTCTGCTATGAAGACACTCAGGAGCAAGATGGGCAAGTGGTGCCGCCA 200

      67 sCysPheProCysCysArgGlySerGlyLysSerAsnValGlyAlaSerG 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     201 CTGCTTCCCCTGCTGCAGGGGGAGTGGCAAGAGCAACGTGGGCGCTTCTG 250

      84 lyAspHisAspAspSerAlaMetLysThrLeuArgAsnLysMetGlyLys 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     251 GAGACCACGACGACTCTGCTATGAAGACACTCAGGAACAAGATGGGCAAG 300

     101 TrpCysCysHisCysPheProCysCysArgGlySerGlyLysSerLysVa 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     301 TGGTGCTGCCACTGCTTCCCCTGCTGCAGGGGGAGCGGCAAGAGCAAGGT 350

     117 lGlyAlaTrpGlyAspTyrAspAspSerAlaPheMetGluProArgTyrH 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     351 GGGCGCTTGGGGAGACTACGATGACAGTGCCTTCATGGAGCCCAGGTACC 400

     134 isValArgGlyGluAspLeuAspLysLeuHisArgAlaAlaTrpTrpGly 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     401 ACGTCCGTGGAGAAGATCTGGACAAGCTCCACAGAGCTGCCTGGTGGGGT 450

     151 LysValProArgLysAspLeuIleValMetLeuArgAspThrAspValAs 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     451 AAAGTCCCCAGAAAGGATCTCATCGTCATGCTCAGGGACACTGACGTGAA 500

     167 nLysLysAspLysGlnLysArgThrAlaLeuHisLeuAlaSerAlaAsnG 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     501 CAAGAAGGACAAGCAAAAGAGGACTGCTCTACATCTGGCCTCTGCCAATG 550

     184 lyAsnSerGluValValLysLeuLeuLeuAspArgArgCysGlnLeuAsn 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     551 GGAATTCAGAAGTAGTAAAACTCCTGCTGGACAGACGATGTCAACTTAAT 600

     201 ValLeuAspAsnLysLysArgThrAlaLeuIleLysAlaValGlnCysGl 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     601 GTCCTTGACAACAAAAAGAGGACAGCTCTGATAAAGGCCGTACAATGCCA 650

     217 nGluAspGluCysAlaLeuMetLeuLeuGluHisGlyThrAspProAsnI 234
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     651 GGAAGATGAATGTGCGTTAATGTTGCTGGAACATGGCACTGATCCAAATA 700

     234 leProAspGluTyrGlyAsnThrThrLeuHisTyrAlaIleTyrAsnGlu 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     701 TTCCAGATGAGTATGGAAATACCACTCTGCACTACGCTATCTATAATGAA 750

     251 AspLysLeuMetAlaLysAlaLeuLeuLeuTyrGlyAlaAspIleGluSe 267
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     751 GATAAATTAATGGCCAAAGCACTGCTCTTATATGGTGCTGATATCGAATC 800

     267 rLysAsnLysHisGlyLeuThrProLeuLeuLeuGlyValHisGluGlnL 284
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     801 AAAAAACAAGCATGGCCTCACACCACTGTTACTTGGTGTACATGAGCAAA 850
```

```
    284 ysGlnGlnValValLysPheLeuIleLysLysLysAlaAsnLeuAsnAla 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    851 AACAGCAAGTCGTGAAATTTTTAATCAAGAAAAAAGCGAATTTAAATGCA 900

    301 LeuAspArgTyrGlyArgThrAlaLeuIleLeuAlaValCysCysGlySe 317
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    901 CTGGATAGATATGGAAGGACTGCTCTCATACTTGCTGTATGTTGTGGATC 950

    317 rAlaSerIleValSerLeuLeuLeuGluGlnAsnIleAspValSerSerG 334
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    951 AGCAAGTATAGTCAGCCTTCTACTTGAGCAAAATATTGATGTATCTTCTC 1000

    334 lnAspLeuSerGlyGlnThrAlaArgGluTyrAlaValSerSerHisHis 350
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1001 AAGATCTATCTGGACAGACGGCCAGAGAGTATGCTGTTTCTAGTCATCAT 1050

    351 HisValIleCysGlnLeuLeuSerAspTyrLysGluLysGlnMetLeuLy 367
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1051 CATGTAATTTGCCAGTTACTTTCTGACTACAAAGAAAAACAGATGCTAAA 1100

    367 sIleSerSerGluAsnSerAsnProGlu 376
        |||||||||||||||||||||||||||
   1101 AATCTCTTCTGAAAACAGCAATCCAGAA 1128

seq_name: gb_pat:AX106594

seq_documentation_block:
LOCUS       AX106594               2040 bp    DNA     linear   PAT 30-APR-2001
DEFINITION  Sequence 375 from Patent WO0125272.
ACCESSION   AX106594
VERSION     AX106594.1  GI:13922265
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2040)
  AUTHORS   Xu,J., Skeiky,Y.A., Reed,S.G. and Cheever,M.A.
  TITLE     Compositions and methods for therapy and diagnosis of prostate
            cancer
  JOURNAL   Patent: WO 0125272-A 375 12-APR-2001;
            CORIXA CORPORATION (US)
FEATURES             Location/Qualifiers
     source          1. .2040
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      716 a    392 c    500 g    432 t
ORIGIN


alignment_scores:
            Quality: 2024.00            Length:    376
              Ratio:  5.383              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-810-936-304 x AX106594   ..

 Align seg 1/1  to: AX106594  from: 1  to: 2040

       1 MetValValGluValAspSerMetProAlaAlaSerServalLysLysPr 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 ATGGTGGTTGAGGTTGATTCCATGCCGGCTGCCTCTTCTGTGAAGAAGCC 50

      17 oPheGlyLeuArgSerLysMetGlyLysTrpCysCysArgCysPheProC 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 ATTTGGTCTCAGGAGCAAGATGGGCAAGTGGTGCTGCCGTTGCTTCCCCT 100

      34 ysCysArgGluSerGlyLysSerAsnValGlyThrSerGlyAspHisAsp 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 GCTGCAGGGAGAGCGGCAAGAGCAACGTGGGCACTTCTGGAGACCACGAC 150

      51 AspSerAlaMetLysThrLeuArgSerLysMetGlyLysTrpCysArgHi 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     151 GACTCTGCTATGAAGACACTCAGGAGCAAGATGGGCAAGTGGTGCCGCCA 200

      67 sCysPheProCysCysArgGlySerGlyLysSerAsnValGlyAlaSerG 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     201 CTGCTTCCCCTGCTGCAGGGGGAGTGGCAAGAGCAACGTGGGCGCTTCTG 250

      84 lyAspHisAspAspSerAlaMetLysThrLeuArgAsnLysMetGlyLys 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     251 GAGACCACGACGACTCTGCTATGAAGACACTCAGGAACAAGATGGGCAAG 300

     101 TrpCysCysHisCysPheProCysCysArgGlySerGlyLysSerLysVa 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     301 TGGTGCTGCCACTGCTTCCCCTGCTGCAGGGGGAGCGGCAAGAGCAAGGT 350

     117 lGlyAlaTrpGlyAspTyrAspAspSerAlaPheMetGluProArgTyrH 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     351 GGGCGCTTGGGGAGACTACGATGACAGTGCCTTCATGGAGCCCAGGTACC 400

     134 isValArgGlyGluAspLeuAspLysLeuHisArgAlaAlaTrpTrpGly 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     401 ACGTCCGTGGAGAAGATCTGGACAAGCTCCACAGAGCTGCCTGGTGGGGT 450

     151 LysValProArgLysAspLeuIleValMetLeuArgAspThrAspValAs 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     451 AAAGTCCCCAGAAAGGATCTCATCGTCATGCTCAGGGACACTGACGTGAA 500

     167 nLysLysAspLysGlnLysArgThrAlaLeuHisLeuAlaSerAlaAsnG 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     501 CAAGAAGGACAAGCAAAAGAGGACTGCTCTACATCTGGCCTCTGCCAATG 550

     184 lyAsnSerGluValValLysLeuLeuLeuAspArgArgCysGlnLeuAsn 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     551 GGAATTCAGAAGTAGTAAAACTCCTGCTGGACAGACGATGTCAACTTAAT 600

     201 ValLeuAspAsnLysLysArgThrAlaLeuIleLysAlaValGlnCysGl 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     601 GTCCTTGACAACAAAAAGAGGACAGCTCTGATAAAGGCCGTACAATGCCA 650

     217 nGluAspGluCysAlaLeuMetLeuLeuGluHisGlyThrAspProAsnI 234
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     651 GGAAGATGAATGTGCGTTAATGTTGCTGGAACATGGCACTGATCCAAATA 700

     234 leProAspGluTyrGlyAsnThrThrLeuHisTyrAlaIleTyrAsnGlu 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     701 TTCCAGATGAGTATGGAAATACCACTCTGCACTACGCTATCTATAATGAA 750

     251 AspLysLeuMetAlaLysAlaLeuLeuLeuTyrGlyAlaAspIleGluSe 267
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     751 GATAAATTAATGGCCAAAGCACTGCTCTTATATGGTGCTGATATCGAATC 800

     267 rLysAsnLysHisGlyLeuThrProLeuLeuLeuGlyValHisGluGlnL 284
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     801 AAAAAACAAGCATGGCCTCACACCACTGTTACTTGGTGTACATGAGCAAA 850

     284 ysGlnGlnValValLysPheLeuIleLysLysLysAlaAsnLeuAsnAla 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     851 AACAGCAAGTCGTGAAATTTTTAATCAAGAAAAAAGCGAATTTAAATGCA 900

     301 LeuAspArgTyrGlyArgThrAlaLeuIleLeuAlaValCysCysGlySe 317
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     901 CTGGATAGATATGGAAGGACTGCTCTCATACTTGCTGTATGTTGTGGATC 950

     317 rAlaSerIleValSerLeuLeuLeuGluGlnAsnIleAspValSerSerG 334
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     951 AGCAAGTATAGTCAGCCTTCTACTTGAGCAAAATATTGATGTATCTTCTC 1000

     334 lnAspLeuSerGlyGlnThrAlaArgGluTyrAlaValSerSerHisHis 350
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1001 AAGATCTATCTGGACAGACGGCCAGAGAGTATGCTGTTTCTAGTCATCAT 1050

     351 HisValIleCysGlnLeuLeuSerAspTyrLysGluLysGlnMetLeuLy 367
```

```
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1051 CATGTAATTTGCCAGTTACTTTCTGACTACAAAGAAAAACAGATGCTAAA 1100

 367 sIleSerSerGluAsnSerAsnProGlu 376
     ||||||||||||||||||||||||||||
1101 AATCTCTTCTGAAAACAGCAATCCAGAA 1128
```

```
OM of: US-09-810-936-304 to: N_Geneseq_032802:*   out_format : pfs

Date: Aug 26, 2002  4:30 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlh
 -Q=/cgn2_1/USPTO_spool/US09810936/runat_23082002_145849_26190/app_query.fasta_1.450
 -DB=N_Geneseq_032802 -QFMT=fastap -SUFFIX=p2n.rng -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
 -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
 -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
 -USER=US09810936_@CGN1_1_174 -NCPU=6 -ICPU=3 -LONGLOG
 -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-810-936-304
Query length: 384
Database: N_Geneseq_032802:*
Database sequences: 1736436
Database length: 858457221
Search time (sec): 203.340000



score_list:
Sequence           Strd Orig    ZScore    EScore Len  ! Documentation  ..
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:AAC81011 + 2064.00 3563.34  3.9e-19
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:AAA06598 + 2064.00 3563.34  3.9e-19
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAI67211 + 2064.00 3563.34  3.9e-1
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAS63807 + 2064.00 3563.34  3.9e-1
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAH93714 + 2064.00 3563.34  3.9e-1
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAH85028 + 2064.00 3563.34  3.9e-1
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAH02779 + 2064.00 3563.34  3.9e-1
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2002.DAT:AAS99857 + 2064.00 3563.34  3.9e-19
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2002.DAT:AAS99869 + 2064.00 3559.87  6.1e-19
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2002.DAT:AAS99872 + 2054.00 3545.97  3.6e-18
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:AAC81012 + 2024.00 3487.91  6.2e-18
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAI67212 + 2024.00 3487.91  6.2e-1
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAS63808 + 2024.00 3487.91  6.2e-1
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAH93715 + 2024.00 3487.91  6.2e-1
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAH85029 + 2024.00 3487.91  6.2e-1
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAH02780 + 2024.00 3487.91  6.2e-1
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2002.DAT:AAS99858 + 2024.00 3487.91  6.2e-18
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:AAC81013 + 2024.00 3487.70  6.4e-18
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:AAA06600 + 2024.00 3487.70  6.4e-18
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAI67213 + 2024.00 3487.70  6.4e-1
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAS63809 + 2024.00 3487.70  6.4e-1
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAH93716 + 2024.00 3487.70  6.4e-1
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAH85030 + 2024.00 3487.70  6.4e-1
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAH02781 + 2024.00 3487.70  6.4e-1
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2002.DAT:AAS99859 + 2024.00 3487.70  6.4e-18
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:AAA06599 + 2020.00 3480.97  1.5e-18
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1998.DAT:AAV68996 + 1920.50 3309.00  5.8e-17
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:AAC81007 + 1920.50 3309.00  5.8e-17
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:AAA06594 + 1920.50 3309.00  5.8e-17
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAS63803 + 1920.50 3309.00  5.8e-1
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAH93710 + 1920.50 3309.00  5.8e-1
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAH85024 + 1920.50 3309.00  5.8e-1
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAH02775 + 1920.50 3309.00  5.8e-1
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2002.DAT:AAS99853 + 1920.50 3309.00  5.8e-17
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2002.DAT:AAS99873 + 1876.00 3236.85  6.0e-17
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2002.DAT:AAS99870 + 1857.00 3203.85  4.1e-17
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2002.DAT:AAS99874 + 1857.00 3203.85  4.1e-17
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:AAC81015 + 1495.00 2570.06  8.3e-13
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAI67210 + 1495.00 2570.06  8.3e-1
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAS63918 + 1495.00 2570.06  8.3e-1
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAH93825 + 1495.00 2570.06  8.3
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAH85139 + 1495.00 2570.06  8.3
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2002.DAT:AAS99861 + 1495.00 2570.06  8.3e
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1998.DAT:AAV68995 + 1469.00 2527.11  2.1e
/SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:AAC81006 + 1469.00 2527.11  2.1e

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:AAC81011

seq_documentation_block:
ID   AAC81011 standard; cDNA; 1155 BP.
XX
AC   AAC81011;
XX
DT   13-FEB-2001  (first entry)
XX
DE   Human B11Ag1 antigen splice isoform B11C-15 cDNA.
XX
KW   Human; breast tumour-specific antigen; cytostatic; vaccine;
KW   breast cancer; B18Ag1; B11Ag1; B15Ag1; ss.
XX
OS   Homo sapiens.
XX
PN   WO200061753-A2.
XX
PD   19-OCT-2000.
XX
PF   07-APR-2000; 2000WO-US09312.
XX
PR   09-APR-1999;   99US-0289198.
PR   28-OCT-1999;   99US-0429755.
PR   23-MAR-2000; 2000US-0534825.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Frudakis TN,  Smith JM,  Reed SG,  Misher LE,  Retter MW,  Dillon DC;
XX
DR   WPI; 2000-628403/60.
DR   P-PSDB; AAB28628.
XX
PT   An isolated polypeptide comprising an immunogenic portion of a breast
PT   tumor protein used for inhibiting the development of cancer, especially
PT   breast cancer, and monitoring cancer progression in a patient -
XX
PS   Claim 4; Page 177; 187pp; English.
XX
CC   The present sequence is given in a specification relating to compositions
CC   and methods for the treatment and diagnosis of breast cancer. Nucleotide
CC   sequences that are preferentially expressed in breast tumour tissue, and
CC   the polypeptides encoded by such nucleotide sequences, are used in
CC   compositions and vaccines to inhibit the development of cancer,
CC   especially breast cancer. The progression of a cancer may be monitored by
CC   carrying out detection of tumour-specific antigens at subsequent time
CC   points and comparing the results from the different time points.
CC   CD4+ and/or CD8+ T-Cells isolated from the cancer patient may be treated
CC   with tumour-specific polypeptides, polynucleotides encoding the
CC   polypeptides or antigen presenting cells expressing the polypeptides. The
CC   cells are then administered to the patient to inhibit development of
CC   cancer.
XX
SQ   Sequence 1155 BP; 346 A; 253 C; 297 G; 259 T; 0 other;


alignment_scores:
            Quality: 2064.00              Length:    384
              Ratio:  5.375                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-810-936-304 x AAC81011   ..

 Align seg 1/1  to: AAC81011  from: 1  to: 1155

       1 MetValValGluValAspSerMetProAlaAlaSerSerValLysLysPr 17
```

```
     1 ATGGTGGTTGAGGTTGATTCCATGCCGGCTGCCTCTTCTGTGAAGAAGCC  50

    17 oPheGlyLeuArgSerLysMetGlyLysTrpCysCysArgCysPheProC  34
       ||||||||||||||||||||||||||||||||||||||||||||||||||
    51 ATTTGGTCTCAGGAGCAAGATGGGCAAGTGGTGCTGCCGTTGCTTCCCCT 100

    34 ysCysArgGluSerGlyLysSerAsnValGlyThrSerGlyAspHisAsp  50
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   101 GCTGCAGGGAGAGCGGCAAGAGCAACGTGGGCACTTCTGGAGACCACGAC 150

    51 AspSerAlaMetLysThrLeuArgSerLysMetGlyLysTrpCysArgHi  67
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   151 GACTCTGCTATGAAGACACTCAGGAGCAAGATGGGCAAGTGGTGCCGCCA 200

    67 sCysPheProCysCysArgGlySerGlyLysSerAsnValGlyAlaSerG  84
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   201 CTGCTTCCCCTGCTGCAGGGGGAGTGGCAAGAGCAACGTGGGCGCTTCTG 250

    84 lyAspHisAspAspSerAlaMetLysThrLeuArgAsnLysMetGlyLys 100
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   251 GAGACCACGACGACTCTGCTATGAAGACACTCAGGAACAAGATGGGCAAG 300

   101 TrpCysCysHisCysPheProCysCysArgGlySerGlyLysSerLysVa 117
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   301 TGGTGCTGCCACTGCTTCCCCTGCTGCAGGGGGAGCGGCAAGAGCAAGGT 350

   117 lGlyAlaTrpGlyAspTyrAspAspSerAlaPheMetGluProArgTyrH 134
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   351 GGGCGCTTGGGGAGACTACGATGACAGTGCCTTCATGGAGCCCAGGTACC 400

   134 isValArgGlyGluAspLeuAspLysLeuHisArgAlaAlaTrpTrpGly 150
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   401 ACGTCCGTGGAGAAGATCTGGACAAGCTCCACAGAGCTGCCTGGTGGGGT 450

   151 LysValProArgLysAspLeuIleValMetLeuArgAspThrAspValAs 167
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   451 AAAGTCCCCAGAAAGGATCTCATCGTCATGCTCAGGGACACTGACGTGAA 500

   167 nLysLysAspLysGlnLysArgThrAlaLeuHisLeuAlaSerAlaAsnG 184
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   501 CAAGAAGGACAAGCAAAAGAGGACTGCTCTACATCTGGCCTCTGCCAATG 550

   184 lyAsnSerGluValValLysLeuLeuLeuAspArgArgCysGlnLeuAsn 200
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   551 GGAATTCAGAAGTAGTAAAACTCCTGCTGGACAGACGATGTCAACTTAAT 600

   201 ValLeuAspAsnLysLysArgThrAlaLeuIleLysAlaValGlnCysGl 217
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   601 GTCCTTGACAACAAAAAGAGGACAGCTCTGATAAAGGCCGTACAATGCCA 650

   217 nGluAspGluCysAlaLeuMetLeuLeuGluHisGlyThrAspProAsnI 234
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   651 GGAAGATGAATGTGCGTTAATGTTGCTGGAACATGGCACTGATCCAAATA 700

   234 leProAspGluTyrGlyAsnThrThrLeuHisTyrAlaIleTyrAsnGlu 250
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   701 TTCCAGATGAGTATGGAAATACCACTCTGCACTACGCTATCTATAATGAA 750

   251 AspLysLeuMetAlaLysAlaLeuLeuLeuTyrGlyAlaAspIleGluSe 267
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   751 GATAAATTAATGGCCAAAGCACTGCTCTTATATGGTGCTGATATCGAATC 800

   267 rLysAsnLysHisGlyLeuThrProLeuLeuLeuGlyValHisGluGlnL 284
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   801 AAAAAACAAGCATGGCCTCACACCACTGTTACTTGGTGTACATGAGCAAA 850

   284 ysGlnGlnValValLysPheLeuIleLysLysLysAlaAsnLeuAsnAla 300
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   851 AACAGCAAGTCGTGAAATTTTTAATCAAGAAAAAAGCGAATTTAAATGCA 900

   301 LeuAspArgTyrGlyArgThrAlaLeuIleLeuAlaValCysCysGlySe 317
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   901 CTGGATAGATATGGAAGGACTGCTCTCATACTTGCTGTATGTTGTGGATC 950

   317 rAlaSerIleValSerLeuLeuLeuGluGlnAsnIleAspValSerSerG 334
       ||||||||||||||||||||||||||||||||||||||||||||||||||
   951 AGCAAGTATAGTCAGCCTTCTACTTGAGCAAAATATTGATGTATCTTCTC 1000

   334 lnAspLeuSerGlyGlnThrAlaArgGluTyrAlaValSerSerHisHis 350
       ||||||||||||||||||||||||||||||||||||||||||||||||||
  1001 AAGATCTATCTGGACAGACGGCCAGAGAGTATGCTGTTTCTAGTCATCAT 1050

   351 HisValIleCysGlnLeuLeuSerAspTyrLysGluLysGlnMetLeuLy 367
       ||||||||||||||||||||||||||||||||||||||||||||||||||
  1051 CATGTAATTTGCCAGTTACTTTCTGACTACAAAGAAAAACAGATGCTAAA 1100

   367 sIleSerSerGluAsnSerAsnProGluAsnValSerArgThrArgAsnL 384
       ||||||||||||||||||||||||||||||||||||||||||||||||||
  1101 AATCTCTTCTGAAAACAGCAATCCAGAAAATGTCTCAAGAACCAGAAATA 1150

   384 ys 384
       ||
  1151 AA 1152
```

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:AAA06598

seq_documentation_block:

ID AAA06598 standard; cDNA; 1155 BP.
XX
AC AAA06598;
XX
DT 13-JUN-2000 (first entry)
XX
DE Human immunogenic prostate tumour protein cDNA sequence SEQ ID NO:373.
XX
KW Human; prostate cancer; diagnosis; tumour; gene therapy; detection;
KW immunogenic; cytostatic; vaccine; ss.
XX
OS Homo sapiens.
XX
PN WO200004149-A2.
XX
PD 27-JAN-2000.
XX
PF 14-JUL-1999; 99WO-US15838.
XX
PR 14-JUL-1998; 98US-0115453.
PR 14-JUL-1998; 98US-0116134.
PR 23-SEP-1998; 98US-0159812.
PR 23-SEP-1998; 98US-0159822.
PR 15-JAN-1999; 99US-0232149.
PR 15-JAN-1999; 99US-0232880.
PR 09-APR-1999; 99US-0288946.
XX
PA (CORI-) CORIXA CORP.
XX
PI Dillon DC, Harlocker SL, Yuqiu J, Xu J, Mitcham JL;
XX
DR WPI; 2000-171268/15.
XX
PT New polypeptide useful for treating and diagnosing prostate cancer
PT comprises an immunogenic portion of prostate tumor protein -
XX
PS Claim 50; Page 222; 263pp; English.
XX
CC The present invention describes isolated polypeptides, comprising an
CC immunogenic portion of a prostate tumour protein (PTP). The polypeptides
CC and polynucleotides encoding them have cytostatic activity and can be
CC used in vaccines and in gene therapy. The polypeptides and
CC polynucleotides encoding them, antigen presenting cells which express
CC the polypeptides, antibodies against the polypeptides and vaccines
CC comprising them can be used for inhibiting the development of prostate
CC cancer in a patient. The polypeptides can be used to generate antibodies
CC or anti-idiotypic antibodies for passive immuno therapy. A portion of
CC the polynucleotides encoding the polypeptides can be used as a probe or

```
CC   to modulate the expression of the polypeptides. AAA06241 to AAA06691 and
CC   AAY82000 to AAY82020 represent sequences used in the exemplification of
CC   the present invention.
XX
SQ   Sequence 1155 BP; 346 A; 253 C; 297 G; 259 T; 0 other;


alignment_scores:
           Quality: 2064.00            Length:    384
             Ratio:  5.375               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-810-936-304 x AAA06598   ..

 Align seg 1/1  to: AAA06598  from: 1  to: 1155

       1 MetValValGluValAspSerMetProAlaAlaSerSerValLysLysPr 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 ATGGTGGTTGAGGTTGATTCCATGCCGGCTGCCTCTTCTGTGAAGAAGCC 50

      17 oPheGlyLeuArgSerLysMetGlyLysTrpCysCysArgCysPheProC 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 ATTTGGTCTCAGGAGCAAGATGGGCAAGTGGTGCTGCCGTTGCTTCCCCT 100

      34 ysCysArgGluSerGlyLysSerAsnValGlyThrSerGlyAspHisAsp 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 GCTGCAGGGAGAGCGGCAAGAGCAACGTGGGCACTTCTGGAGACCACGAC 150

      51 AspSerAlaMetLysThrLeuArgSerLysMetGlyLysTrpCysArgHi 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     151 GACTCTGCTATGAAGACACTCAGGAGCAAGATGGGCAAGTGGTGCCGCCA 200

      67 sCysPheProCysCysArgGlySerGlyLysSerAsnValGlyAlaSerG 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     201 CTGCTTCCCCTGCTGCAGGGGGAGTGGCAAGAGCAACGTGGGCGCTTCTG 250

      84 lyAspHisAspAspSerAlaMetLysThrLeuArgAsnLysMetGlyLys 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     251 GAGACCACGACGACTCTGCTATGAAGACACTCAGGAACAAGATGGGCAAG 300

     101 TrpCysCysHisCysPheProCysCysArgGlySerGlyLysSerLysVa 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     301 TGGTGCTGCCACTGCTTCCCCTGCTGCAGGGGGAGCGGCAAGAGCAAGGT 350

     117 lGlyAlaTrpGlyAspTyrAspAspSerAlaPheMetGluProArgTyrH 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     351 GGGCGCTTGGGGAGACTACGATGACAGTGCCTTCATGGAGCCCAGGTACC 400

     134 isValArgGlyGluAspLeuAspLysLeuHisArgAlaAlaTrpTrpGly 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     401 ACGTCCGTGGAGAAGATCTGGACAAGCTCCACAGAGCTGCCTGGTGGGGT 450

     151 LysValProArgLysAspLeuIleValMetLeuArgAspThrAspValAs 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     451 AAAGTCCCCAGAAAGGATCTCATCGTCATGCTCAGGGACACTGACGTGAA 500

     167 nLysLysAspLysGlnLysArgThrAlaLeuHisLeuAlaSerAlaAsnG 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     501 CAAGAAGGACAAGCAAAAGAGGACTGCTCTACATCTGGCCTCTGCCAATG 550

     184 lyAsnSerGluValValLysLeuLeuLeuAspArgArgCysGlnLeuAsn 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     551 GGAATTCAGAAGTAGTAAAACTCCTGCTGGACAGACGATGTCAACTTAAT 600

     201 ValLeuAspAsnLysLysArgThrAlaLeuIleLysAlaValGlnCysGl 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     601 GTCCTTGACAACAAAAAGAGGACAGCTCTGATAAAGGCCGTACAATGCCA 650

     217 nGluAspGluCysAlaLeuMetLeuLeuGluHisGlyThrAspProAsnI 234
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     651 GGAAGATGAATGTGCGTTAATGTTGCTGGAACATGGCACTGATCCAAATA 700

     234 leProAspGluTyrGlyAsnThrThrLeuHisTyrAlaIleTyrAsnGlu 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     701 TTCCAGATGAGTATGGAAATACCACTCTGCACTACGCTATCTATAATGAA 750

     251 AspLysLeuMetAlaLysAlaLeuLeuLeuTyrGlyAlaAspIleGluSe 267
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     751 GATAAATTAATGGCCAAAGCACTGCTCTTATATGGTGCTGATATCGAATC 800

     267 rLysAsnLysHisGlyLeuThrProLeuLeuLeuGlyValHisGluGlnL 284
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     801 AAAAAACAAGCATGGCCTCACACCACTGTTACTTGGTGTACATGAGCAAA 850

     284 ysGlnGlnValValLysPheLeuIleLysLysLysAlaAsnLeuAsnAla 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     851 AACAGCAAGTCGTGAAATTTTTAATCAAGAAAAAAGCGAATTTAAATGCA 900

     301 LeuAspArgTyrGlyArgThrAlaLeuIleLeuAlaValCysCysGlySe 317
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     901 CTGGATAGATATGGAAGGACTGCTCTCATACTTGCTGTATGTTGTGGATC 950

     317 rAlaSerIleValSerLeuLeuLeuGluGlnAsnIleAspValSerSerG 334
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     951 AGCAAGTATAGTCAGCCTTCTACTTGAGCAAAATATTGATGTATCTTCTC 1000

     334 lnAspLeuSerGlyGlnThrAlaArgGluTyrAlaValSerSerHisHis 350
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1001 AAGATCTATCTGGACAGACGGCCAGAGAGTATGCTGTTTCTAGTCATCAT 1050

     351 HisValIleCysGlnLeuLeuSerAspTyrLysGluLysGlnMetLeuLy 367
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1051 CATGTAATTTGCCAGTTACTTTCTGACTACAAAGAAAAACAGATGCTAAA 1100

     367 sIleSerSerGluAsnSerAsnProGluAsnValSerArgThrArgAsnL 384
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1101 AATCTCTTCTGAAAACAGCAATCCAGAAAATGTCTCAAGAACCAGAAATA 1150

     384 ys 384
         ||
    1151 AA 1152

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAI67211

seq_documentation_block:
ID   AAI67211 standard; cDNA; 1155 BP.
XX
AC   AAI67211;
XX
DT   11-FEB-2002  (first entry)
XX
DE   B305D isoform C splice variant 1 encoding cDNA.
XX
KW   Genetic subtraction; DNA microarray analysis; polymerase chain reaction;
KW   cancer; B305D; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1155
FT                   /*tag=  a
FT                   /product= "B305D isoform C splice variant"
XX
PN   WO200175171-A2.
XX
PD   11-OCT-2001.
XX
PF   02-APR-2001; 2001WO-US10631.
XX
PR   03-APR-2000; 2000US-194241P.
PR   20-JUL-2000; 2000US-219862P.
PR   27-JUL-2000; 2000US-221300P.
PR   18-DEC-2000; 2000US-256592P.
XX
PA   (CORI-) CORIXA CORP.
```

```
XX
PI   Houghton RL,  Dillon DC,  Molesh DA,  Xu J,  Zehentner B,  Persing DH;
XX
DR   WPI; 2001-626449/72.
DR   P-PSDB; AAG65976.
XX
PT   Identifying tissue (tumour)-specific polynucleotides overexpressed in
PT   tissue of interest as compared to control tissue, for detecting cancer
PT   cells in patient, comprises DNA microarray analysis or quantitative
PT   polymerase chain reaction -
XX
PS   Claim 4; Page 93-94; 127pp; English.
XX
CC   The invention relates to identifying tissue-specific polynucleotides (P)
CC   that involves performing a genetic subtraction to identify pool of (P)
CC   from tissue of interest (TI), performing DNA microarray analysis to
CC   identify first subset of polynucleotides (SP1) at least 2-fold over
CC   expressed in TI, and performing quantitative polymerase chain reaction
CC   (PCR) analysis on SP1 to identify second subset of (P). The method is
CC   useful for determining the presence or absence of a cancer cell in a
CC   patient, monitoring the progression of cancer in a patient using a
CC   biological sample such as blood, serum, lymph nodes, bone marrow, sputum,
CC   urine or a tumour biopsy sample. The methods are useful for determining
CC   the presence or absence of or monitoring progression of prostate, breast,
CC   colon, ovarian, lung, head and neck, lymphoma, leukemia, melanoma, liver,
CC   gastric, kidney, bladder, pancreatic or endometrial cancer. The present
CC   sequence represents a cDNA encoding a B305D isoform C splice variant.
XX
SQ   Sequence 1155 BP; 346 A; 253 C; 297 G; 259 T; 0 other;


alignment_scores:
            Quality: 2064.00              Length:    384
              Ratio:  5.375                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-810-936-304 x AAI67211   ..

 Align seg 1/1  to: AAI67211  from: 1  to: 1155

       1 MetValValGluValAspSerMetProAlaAlaSerSerValLysLysPr 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 ATGGTGGTTGAGGTTGATTCCATGCCGGCTGCCTCTTCTGTGAAGAAGCC 50

      17 oPheGlyLeuArgSerLysMetGlyLysTrpCysCysArgCysPheProC 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 ATTTGGTCTCAGGAGCAAGATGGGCAAGTGGTGCTGCCGTTGCTTCCCCT 100

      34 ysCysArgGluSerGlyLysSerAsnValGlyThrSerGlyAspHisAsp 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 GCTGCAGGGAGAGCGGCAAGAGCAACGTGGGCACTTCTGGAGACCACGAC 150

      51 AspSerAlaMetLysThrLeuArgSerLysMetGlyLysTrpCysArgHi 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     151 GACTCTGCTATGAAGACACTCAGGAGCAAGATGGGCAAGTGGTGCCGCCA 200

      67 sCysPheProCysCysArgGlySerGlyLysSerAsnValGlyAlaSerG 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     201 CTGCTTCCCCTGCTGCAGGGGGAGTGGCAAGAGCAACGTGGGCGCTTCTG 250

      84 lyAspHisAspSerAlaMetLysThrLeuArgAsnLysMetGlyLys 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     251 GAGACCACGACGACTCTGCTATGAAGACACTCAGGAACAAGATGGGCAAG 300

     101 TrpCysCysHisCysPheProCysCysArgGlySerGlyLysSerLysVa 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     301 TGGTGCTGCCACTGCTTCCCCTGCTGCAGGGGGAGCGGCAAGAGCAAGGT 350

     117 lGlyAlaTrpGlyAspTyrAspAspSerAlaPheMetGluProArgTyrH 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     351 GGGCGCTTGGGGAGACTACGATGACAGTGCCTTCATGGAGCCCAGGTACC 400

     134 isValArgGlyGluAspLeuAspLysLeuHisArgAlaAlaTrpTrpGly 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     401 ACGTCCGTGGAGAAGATCTGGACAAGCTCCACAGAGCTGCCTGGTGGGGT 450

     151 LysValProArgLysAspLeuIleValMetLeuArgAspThrAspValAs 167
         |||||||||||||||||||||||||||||||||||||||||||||||||
     451 AAAGTCCCCAGAAAGGATCTCATCGTCATGCTCAGGGACACTGACGTGAA 500

     167 nLysLysAspLysGlnLysArgThrAlaLeuHisLeuAlaSerAlaAsnG 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     501 CAAGAAGGACAAGCAAAAGAGGACTGCTCTACATCTGGCCTCTGCCAATG 550

     184 lyAsnSerGluValValLysLeuLeuLeuAspArgArgCysGlnLeuAsn 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     551 GGAATTCAGAAGTAGTAAAACTCCTGCTGGACAGACGATGTCAACTTAAT 600

     201 ValLeuAspAsnLysLysArgThrAlaLeuIleLysAlaValGlnCysGl 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     601 GTCCTTGACAACAAAAAGAGGACAGCTCTGATAAAGGCCGTACAATGCCA 650

     217 nGluAspGluCysAlaLeuMetLeuLeuGluHisGlyThrAspProAsnI 234
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     651 GGAAGATGAATGTGCGTTAATGTTGCTGGAACATGGCACTGATCCAAATA 700

     234 leProAspGluTyrGlyAsnThrThrLeuHisTyrAlaIleTyrAsnGlu 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     701 TTCCAGATGAGTATGGAAATACCACTCTGCACTACGCTATCTATAATGAA 750

     251 AspLysLeuMetAlaLysAlaLeuLeuLeuTyrGlyAlaAspIleGluSe 267
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     751 GATAAATTAATGGCCAAAGCACTGCTCTTATATGGTGCTGATATCGAATC 800

     267 rLysAsnLysHisGlyLeuThrProLeuLeuLeuGlyValHisGluGlnL 284
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     801 AAAAAACAAGCATGGCCTCACACCACTGTTACTTGGTGTACATGAGCAAA 850

     284 ysGlnGlnValValLysPheLeuIleLysLysLysAlaAsnLeuAsnAla 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     851 AACAGCAAGTCGTGAAATTTTTAATCAAGAAAAAAGCGAATTTAAATGCA 900

     301 LeuAspArgTyrGlyArgThrAlaLeuIleLeuAlaValCysCysGlySe 317
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     901 CTGGATAGATATGGAAGGACTGCTCTCATACTTGCTGTATGTTGTGGATC 950

     317 rAlaSerIleValSerLeuLeuLeuGluGlnAsnIleAspValSerSerG 334
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     951 AGCAAGTATAGTCAGCCTTCTACTTGAGCAAAATATTGATGTATCTTCTC 1000

     334 lnAspLeuSerGlyGlnThrAlaArgGluTyrAlaValSerSerHisHis 350
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1001 AAGATCTATCTGGACAGACGGCCAGAGAGTATGCTGTTTCTAGTCATCAT 1050

     351 HisValIleCysGlnLeuLeuSerAspTyrLysGluLysGlnMetLeuLy 367
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1051 CATGTAATTTGCCAGTTACTTTCTGACTACAAAGAAAAACAGATGCTAAA 1100

     367 sIleSerSerGluAsnSerAsnProGluAsnValSerArgThrArgAsnL 384
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1101 AATCTCTTCTGAAAACAGCAATCCAGAAAATGTCTCAAGAACCAGAAATA 1150

     384 ys 384
         ||
    1151 AA 1152

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAS63807

seq_documentation_block:
ID   AAS63807 standard; cDNA; 1155 BP.
XX
AC   AAS63807;
XX
DT   29-JAN-2002  (first entry)
XX
```

```
DE   Human prostate cDNA clone B305D splice variant #8.
XX
KW   Human; prostate cancer; ss; cytostatic; immunostimulant; tumour.
XX
OS   Homo sapiens.
XX
PN   WO200173032-A2.
XX
PD   04-OCT-2001.
XX
PF   27-MAR-2001; 2001WO-US09919.
XX
PR   27-MAR-2000; 2000US-0536857.
PR   09-MAY-2000; 2000US-0568100.
PR   12-MAY-2000; 2000US-0570737.
PR   13-JUN-2000; 2000US-0593793.
PR   27-JUN-2000; 2000US-0605783.
PR   10-AUG-2000; 2000US-0636215.
PR   29-AUG-2000; 2000US-0651236.
PR   06-SEP-2000; 2000US-0657279.
PR   02-OCT-2000; 2000US-0679426.
PR   10-OCT-2000; 2000US-0685166.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD;
PI   Fanger GR, Retter MW, Stolk JA, Day CH, Vedvick TS, Carter D;
PI   Li SX, Wang A, Skeiky YAW, Hepler WT, Henderson RA;
XX
DR   WPI; 2001-639232/73.
DR   P-PSDB; AAU69777.
XX
PT   New human prostate-specific polypeptides and polynucleotides useful for
PT   the diagnosis and treatment of cancer, especially prostate cancer -
XX
PS   Claim 1; Page 349; 579pp; English.
XX
CC   The invention relates to isolated prostate-specific
CC   polynucleotides, polypeptides, fusion proteins of the polypeptides,
CC   antibodies raised against the polypeptides (or antigenic epitopes
CC   derived from them) and antigen-presenting cells expressing the
CC   polypeptides. The antibodies are useful for detecting the presence of
CC   cancer, especially prostate cancer. The polypeptides, polynucleotides and
CC   the antigen-presenting cells are useful for stimulating and/or expanding
CC   T cells specific for a tumour protein, and for inhibiting the development
CC   of cancer especially prostate cancer. Compositions comprising the
CC   polynucleotide and/or polypeptide are useful for stimulating an immune
CC   response, and for treating cancer. The oligonucleotide is useful for
CC   detecting cancer. The present sequence is a prostate specific
CC   polynucleotide of the invention.
XX
SQ   Sequence 1155 BP; 346 A; 253 C; 297 G; 259 T; 0 other;


alignment_scores:
            Quality: 2064.00              Length:    384
              Ratio:  5.375                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-810-936-304 x AAS63807   ..

 Align seg 1/1  to: AAS63807  from: 1  to: 1155

       1 MetValValGluValAspSerMetProAlaAlaSerSerValLysLysPr 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 ATGGTGGTTGAGGTTGATTCCATGCCGGCTGCCTCTTCTGTGAAGAAGCC 50

      17 oPheGlyLeuArgSerLysMetGlyLysTrpCysCysArgCysPheProC 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 ATTTGGTCTCAGGAGCAAGATGGGCAAGTGGTGCTGCCGTTGCTTCCCCT 100

      34 ysCysArgGluSerGlyLysSerAsnValGlyThrSerGlyAspHisAsp 50
                                                              ||||||||||||||||||||||||||||||||||||||||||||||||||
                                                            101 GCTGCAGGGAGAGCGGCAAGAGCAACGTGGGCACTTCTGGAGACCACGAC 150

                                                             51 AspSerAlaMetLysThrLeuArgSerLysMetGlyLysTrpCysArgHi 67
                                                                ||||||||||||||||||||||||||||||||||||||||||||||||||
                                                            151 GACTCTGCTATGAAGACACTCAGGAGCAAGATGGGCAAGTGGTGCCGCCA 200

                                                             67 sCysProCysCysArgGlySerGlyLysSerAsnValGlyAlaSerG 84
                                                                ||||||||||||||||||||||||||||||||||||||||||||||||
                                                            201 CTGCTTCCCCTGCTGCAGGGGGAGTGGCAAGAGCAACGTGGGCGCTTCTG 250

                                                             84 lyAspHisAspAspSerAlaMetLysThrLeuArgAsnLysMetGlyLys 100
                                                                ||||||||||||||||||||||||||||||||||||||||||||||||
                                                            251 GAGACCACGACGACTCTGCTATGAAGACACTCAGGAACAAGATGGGCAAG 300

                                                            101 TrpCysCysHisCysProCysCysArgGlySerGlyLysSerLysVa 117
                                                                ||||||||||||||||||||||||||||||||||||||||||||||||
                                                            301 TGGTGCTGCCACTGCTTCCCCTGCTGCAGGGGGAGCGGCAAGAGCAAGGT 350

                                                            117 lGlyAlaTrpGlyAspTyrAspAspSerAlaPheMetGluProArgTyrH 134
                                                                ||||||||||||||||||||||||||||||||||||||||||||||||
                                                            351 GGGCGCTTGGGGAGACTACGATGACAGTGCCTTCATGGAGCCCAGGTACC 400

                                                            134 isValArgGlyGluAspLeuAspLysLeuHisArgAlaAlaTrpTrpGly 150
                                                                ||||||||||||||||||||||||||||||||||||||||||||||||
                                                            401 ACGTCCGTGGAGAAGATCTGGACAAGCTCCACAGAGCTGCCTGGTGGGGT 450

                                                            151 LysValProArgLysAspLeuIleValMetLeuArgAspThrAspValAs 167
                                                                ||||||||||||||||||||||||||||||||||||||||||||||||
                                                            451 AAAGTCCCCAGAAAGGATCTCATCGTCATGCTCAGGGACACTGACGTGAA 500

                                                            167 nLysLysAspLysGlnLysArgThrAlaLeuHisLeuAlaSerAlaAsnG 184
                                                                ||||||||||||||||||||||||||||||||||||||||||||||||
                                                            501 CAAGAAGGACAAGCAAAAGAGGACTGCTCTACATCTGGCCTCTGCCAATG 550

                                                            184 lyAsnSerGluValValLysLeuLeuLeuAspArgArgCysGlnLeuAsn 200
                                                                ||||||||||||||||||||||||||||||||||||||||||||||||
                                                            551 GGAATTCAGAAGTAGTAAAACTCCTGCTGGACAGACGATGTCAACTTAAT 600

                                                            201 ValLeuAspAsnLysLysArgThrAlaLeuIleLysAlaValGlnCysGl 217
                                                                ||||||||||||||||||||||||||||||||||||||||||||||||
                                                            601 GTCCTTGACAACAAAAAGAGGACAGCTCTGATAAAGGCCGTACAATGCCA 650

                                                            217 nGluAspGluCysAlaLeuMetLeuLeuGluHisGlyThrAspProAsnI 234
                                                                ||||||||||||||||||||||||||||||||||||||||||||||||
                                                            651 GGAAGATGAATGTGCGTTAATGTTGCTGGAACATGGCACTGATCCAAATA 700

                                                            234 leProAspGluTyrGlyAsnThrThrLeuHisTyrAlaIleTyrAsnGlu 250
                                                                ||||||||||||||||||||||||||||||||||||||||||||||||
                                                            701 TTCCAGATGAGTATGGAAATACCACTCTGCACTACGCTATCTATAATGAA 750

                                                            251 AspLysLeuMetAlaLysAlaLeuLeuLeuTyrGlyAlaAspIleGluSe 267
                                                                ||||||||||||||||||||||||||||||||||||||||||||||||
                                                            751 GATAAATTAATGGCCAAAGCACTGCTCTTATATGGTGCTGATATCGAATC 800

                                                            267 rLysAsnLysHisGlyLeuThrProLeuLeuLeuGlyValHisGluGlnL 284
                                                                ||||||||||||||||||||||||||||||||||||||||||||||||
                                                            801 AAAAAACAAGCATGGCCTCACACCACTGTTACTTGGTGTACATGAGCAAA 850

                                                            284 ysGlnGlnValValLysPheLeuIleLysLysLysAlaAsnLeuAsnAla 300
                                                                ||||||||||||||||||||||||||||||||||||||||||||||||
                                                            851 AACAGCAAGTCGTGAAATTTTTAATCAAGAAAAAAGCGAATTTAAATGCA 900

                                                            301 LeuAspArgTyrGlyArgThrAlaLeuIleLeuAlaValCysCysGlySe 317
                                                                ||||||||||||||||||||||||||||||||||||||||||||||||
                                                            901 CTGGATAGATATGGAAGGACTGCTCTCATACTTGCTGTATGTTGTGGATC 950

                                                            317 rAlaSerIleValSerLeuLeuLeuGluGlnAsnIleAspValSerSerG 334
                                                                ||||||||||||||||||||||||||||||||||||||||||||||||
                                                            951 AGCAAGTATAGTCAGCCTTCTACTTGAGCAAAATATTGATGTATCTTCTC 1000

                                                            334 lnAspLeuSerGlyGlnThrAlaArgGluTyrAlaValSerSerHisHis 350
                                                                ||||||||||||||||||||||||||||||||||||||||||||||||
```

```
    1001 AAGATCTATCTGGACAGACGGCCAGAGAGTATGCTGTTTCTAGTCATCAT 1050

     351 HisValIleCysGlnLeuLeuSerAspTyrLysGluLysGlnMetLeuLy 367
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    1051 CATGTAATTTGCCAGTTACTTTCTGACTACAAAGAAAAACAGATGCTAAA 1100

     367 sIleSerSerGluAsnSerAsnProGluAsnValSerArgThrArgAsnL 384
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    1101 AATCTCTTCTGAAAACAGCAATCCAGAAAATGTCTCAAGAACCAGAAATA 1150

     384 ys 384
         ||
    1151 AA 1152

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAH93714

seq_documentation_block:
ID   AAH93714 standard; cDNA; 1155 BP.
XX
AC   AAH93714;
XX
DT   04-OCT-2001  (first entry)
XX
DE   Human prostate-specific cDNA sequence B305D splice variant #8.
XX
KW   Human; prostate cancer; prostate-specific; diagnosis; vaccine;
KW   cytostatic; gene therapy; metastasis; ss.
XX
OS   Homo sapiens.
XX
PN   WO200151633-A2.
XX
PD   19-JUL-2001.
XX
PF   16-JAN-2001; 2001WO-US01574.
XX
PR   14-JAN-2000; 2000US-0483672.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Xu J,  Dillon DC,  Mitcham JL,  Harlocker SL,  Jiang Y,  Reed SG;
PI   Kalos MD,  Fanger GR,  Day CH,  Retter MW,  Stolk JA,  Skeiky YAW;
PI   Wang A,  Meagher MJ;
XX
DR   WPI; 2001-425873/45.
XX
PT   New polynucleotide encoding a prostate-specific protein, for
PT   diagnosing, monitoring and treating prostate cancer in a patient and
PT   for use in vaccines -
XX
PS   Claim 1; Page 347; 543pp; English.
XX
CC   The present invention describes polynucleotide sequences (I) which encode
CC   prostate-specific proteins (II). (I) and (II) have cytostatic activity,
CC   and can be used in vaccine production and gene therapy. (I), (II),
CC   antibodies to (II), fusion proteins comprising (II), and isolated
CC   T cells prepared using (I) or (II) are used treat cancer in a patient.
CC   (I) and the antibodies are also used in the detection of cancer in a
CC   patient. The cancer that is diagnosed or treated is particularly
CC   prostate cancer. (I) and (II) can be used in vaccines. The antibodies or
CC   (I) can be used for monitoring the progression of cancer in a patient.
CC   (I) and (II) can also be used to improve diagnostic and therapeutic
CC   methods for prostate cancer. They can indicate the level of metastasis
CC   as well as the prostate volume. AAH93357 to AAH93944 and AAM01115 to
CC   AAM01318 represent polynucleotide and amino acid sequences used in the
CC   exemplification of the present invention.
XX
SQ   Sequence 1155 BP; 346 A; 253 C; 297 G; 259 T; 0 other;


alignment_scores:
            Quality: 2064.00            Length:    384
              Ratio:  5.375              Gaps:      0

 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-810-936-304 x AAH93714   ..

 Align seg 1/1  to: AAH93714  from: 1  to: 1155

       1 MetValValGluValAspSerMetProAlaAlaSerSerValLysLysPr 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 ATGGTGGTTGAGGTTGATTCCATGCCGGCTGCCTCTTCTGTGAAGAAGCC 50

      17 oPheGlyLeuArgSerLysMetGlyLysTrpCysCysArgCysPheProC 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 ATTTGGTCTCAGGAGCAAGATGGGCAAGTGGTGCTGCCGTTGCTTCCCCT 100

      34 ysCysArgGluSerGlyLysSerAsnValGlyThrSerGlyAspHisAsp 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 GCTGCAGGGAGAGCGGCAAGAGCAACGTGGGCACTTCTGGAGACCACGAC 150

      51 AspSerAlaMetLysThrLeuArgSerLysMetGlyLysTrpCysArgHi 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     151 GACTCTGCTATGAAGACACTCAGGAGCAAGATGGGCAAGTGGTGCCGCCA 200

      67 sCysPheProCysCysArgGlySerGlyLysSerAsnValGlyAlaSerG 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     201 CTGCTTCCCCTGCTGCAGGGGGAGTGGCAAGAGCAACGTGGGCGCTTCTG 250

      84 lyAspHisAspAspSerAlaMetLysThrLeuArgAsnLysMetGlyLys 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     251 GAGACCACGACGACTCTGCTATGAAGACACTCAGGAACAAGATGGGCAAG 300

     101 TrpCysCysHisCysPheProCysCysArgGlySerGlyLysSerLysVa 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     301 TGGTGCTGCCACTGCTTCCCCTGCTGCAGGGGGAGCGGCAAGAGCAAGGT 350

     117 lGlyAlaTrpGlyAspTyrAspAspSerAlaPheMetGluProArgTyrH 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     351 GGGCGCTTGGGGAGACTACGATGACAGTGCCTTCATGGAGCCCAGGTACC 400

     134 isValArgGlyGluAspLeuAspLysLeuHisArgAlaAlaTrpTrpGly 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     401 ACGTCCGTGGAGAAGATCTGGACAAGCTCCACAGAGCTGCCTGGTGGGGT 450

     151 LysValProArgLysAspLeuIleValMetLeuArgAspThrAspValAs 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     451 AAAGTCCCCAGAAAGGATCTCATCGTCATGCTCAGGGACACTGACGTGAA 500

     167 nLysLysAspLysGlnLysArgThrAlaLeuHisLeuAlaSerAlaAsnG 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     501 CAAGAAGGACAAGCAAAAGAGGACTGCTCTACATCTGGCCTCTGCCAATG 550

     184 lyAsnSerGluValValLysLeuLeuLeuAspArgArgCysGlnLeuAsn 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     551 GGAATTCAGAAGTAGTAAAACTCCTGCTGGACAGACGATGTCAACTTAAT 600

     201 ValLeuAspAsnLysLysArgThrAlaLeuIleLysAlaValGlnCysGl 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     601 GTCCTTGACAACAAAAAGAGGACAGCTCTGATAAAGGCCGTACAATGCCA 650

     217 nGluAspGluCysAlaLeuMetLeuLeuGluHisGlyThrAspProAsnI 234
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     651 GGAAGATGAATGTGCGTTAATGTTGCTGGAACATGGCACTGATCCAAATA 700

     234 leProAspGluTyrGlyAsnThrThrLeuHisTyrAlaIleTyrAsnGlu 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     701 TTCCAGATGAGTATGGAAATACCACTCTGCACTACGCTATCTATAATGAA 750

     251 AspLysLeuMetAlaLysAlaLeuLeuLeuTyrGlyAlaAspIleGluSe 267
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     751 GATAAATTAATGGCCAAAGCACTGCTCTTATATGGTGCTGATATCGAATC 800

     267 rLysAsnLysHisGlyLeuThrProLeuLeuLeuGlyValHisGluGlnL 284
         ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
     801 AAAAAACAAGCATGGCCTCACACCACTGTTACTTGGTGTACATGAGCAAA 850

     284 ysGlnGlnValValLysPheLeuIleLysLysLysAlaAsnLeuAsnAla 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     851 AACAGCAAGTCGTGAAATTTTTAATCAAGAAAAAAGCGAATTTAAATGCA 900

     301 LeuAspArgTyrGlyArgThrAlaLeuIleLeuAlaValCysCysGlySe 317
         |||||||||||||||||||||||||||||||||||||||||||||||||
     901 CTGGATAGATATGGAAGGACTGCTCTCATACTTGCTGTATGTTGTGGATC 950

     317 rAlaSerIleValSerLeuLeuLeuGluGlnAsnIleAspValSerSerG 334
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     951 AGCAAGTATAGTCAGCCTTCTACTTGAGCAAAATATTGATGTATCTTCTC 1000

     334 lnAspLeuSerGlyGlnThrAlaArgGluTyrAlaValSerSerHisHis 350
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1001 AAGATCTATCTGGACAGACGGCCAGAGAGTATGCTGTTTCTAGTCATCAT 1050

     351 HisValIleCysGlnLeuLeuSerAspTyrLysGluLysGlnMetLeuLy 367
         |||||||||||||||||||||||||||||||||||||||||||||||||
    1051 CATGTAATTTGCCAGTTACTTTCTGACTACAAAGAAAAACAGATGCTAAA 1100

     367 sIleSerSerGluAsnSerAsnProGluAsnValSerArgThrArgAsnL 384
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1101 AATCTCTTCTGAAAACAGCAATCCAGAAAATGTCTCAAGAACCAGAAATA 1150

     384 ys 384
         ||
    1151 AA 1152

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAH85028

seq_documentation_block:
ID   AAH85028 standard; cDNA; 1155 BP.
XX
AC   AAH85028;
XX
DT   25-SEP-2001  (first entry)
XX
DE   Human prostate-specific cDNA sequence B305D splice variant #8.
XX
KW   Human; prostate cancer; therapy; diagnosis; cat eye syndrome;
KW   chromosome 22q11.2; prostate-specific protein; chromosome 1;
KW   prostate specific antigen; PSA; ss.
XX
OS   Homo sapiens.
XX
PN   WO200134802-A2.
XX
PD   17-MAY-2001.
XX
PF   09-NOV-2000; 2000WO-US30904.
XX
PR   12-NOV-1999;   99US-0439313.
PR   18-NOV-1999;   99US-0443686.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Xu J,  Dillon DC,  Mitcham JL,  Harlocker SL,  Jiang Y,  Reed SG;
PI   Kalos MD,  Retter MW,  Stolk JA,  Day CH,  Skeiky YAW,  Wang A;
XX
DR   WPI; 2001-308785/32.
XX
PT   Isolated polypeptide comprising at least an immunogenic portion of a
PT   prostate-specific protein, useful in the diagnosis and therapy of
PT   prostate cancer -
XX
PS   Claim 31; Page 246-247; 325pp; English.
XX
CC   The present invention describes an isolated polypeptide (P1) comprising
CC   at least an immunogenic portion of a prostate-specific protein, or its
CC   variant. Also described are polynucleotides (N1) encoding (P1). (P1) and
CC   (N1) have cytostatic activity and can be used in vaccine production.
CC   The polypeptides, nucleic acids and antibodies from the present
CC   invention are useful in the diagnosis and therapy of prostate cancer.
CC   Prostate specific genes P704P, P712P, P774P, P775P and B305D are located
CC   in a genomic region on chromosome 22q11.2 known as the Cat Eye Syndrome
CC   region. Prostate specific antigen (PSA) P501S was located on
CC   chromosome 1. AAH84671 to AAH85143 and AAG99000 to AAG99077 represent
CC   polynucleotide and polypeptide sequences used in the exemplification
CC   of the present invention.
XX
SQ   Sequence 1155 BP; 346 A; 253 C; 297 G; 259 T; 0 other;


alignment_scores:
            Quality: 2064.00              Length:    384
              Ratio:  5.375                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-810-936-304 x AAH85028   ..

 Align seg 1/1  to: AAH85028  from: 1  to: 1155

       1 MetValValGluValAspSerMetProAlaAlaSerSerValLysLysPr 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 ATGGTGGTTGAGGTTGATTCCATGCCGGCTGCCTCTTCTGTGAAGAAGCC 50

      17 oPheGlyLeuArgSerLysMetGlyLysTrpCysCysArgCysPheProC 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 ATTTGGTCTCAGGAGCAAGATGGGCAAGTGGTGCTGCCGTTGCTTCCCCT 100

      34 ysCysArgGluSerGlyLysSerAsnValGlyThrSerGlyAspHisAsp 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 GCTGCAGGGAGAGCGGCAAGAGCAACGTGGGCACTTCTGGAGACCACGAC 150

      51 AspSerAlaMetLysThrLeuArgSerLysMetGlyLysTrpCysArgHi 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     151 GACTCTGCTATGAAGACACTCAGGAGCAAGATGGGCAAGTGGTGCCGCCA 200

      67 sCysProCysCysArgGlySerGlyLysSerAsnValGlyAlaSerG 84
         ||||||||||||||||||||||||||||||||||||||||||||||||
     201 CTGCTTCCCCTGCTGCAGGGGGAGTGGCAAGAGCAACGTGGGCGCTTCTG 250

      84 lyAspHisAspAspSerAlaMetLysThrLeuArgAsnLysMetGlyLys 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     251 GAGACCACGACGACTCTGCTATGAAGACACTCAGGAACAAGATGGGCAAG 300

     101 TrpCysCysHisCysProProCysCysArgGlySerGlyLysSerLysVa 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     301 TGGTGCTGCCACTGCTTCCCCTGCTGCAGGGGGAGCGGCAAGAGCAAGGT 350

     117 lGlyAlaTrpGlyAspTyrAspAspSerAlaPheMetGluProArgTyrH 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     351 GGGCGCTTGGGGAGACTACGATGACAGTGCCTTCATGGAGCCCAGGTACC 400

     134 isValArgGlyGluAspLeuAspLysLeuHisArgAlaAlaTrpTrpGly 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     401 ACGTCCGTGGAGAAGATCTGGACAAGCTCCACAGAGCTGCCTGGTGGGGT 450

     151 LysValProArgLysAspLeuIleValMetLeuArgAspThrAspValAs 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     451 AAAGTCCCCAGAAAGGATCTCATCGTCATGCTCAGGGACACTGACGTGAA 500

     167 nLysLysAspLysGlnLysArgThrAlaLeuHisLeuAlaSerAlaAsnG 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     501 CAAGAAGGACAAGCAAAAGAGGACTGCTCTACATCTGGCCTCTGCCAATG 550

     184 lyAsnSerGluValValLysLeuLeuLeuAspArgArgCysGlnLeuAsn 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     551 GGAATTCAGAAGTAGTAAAACTCCTGCTGGACAGACGATGTCAACTTAAT 600

     201 ValLeuAspAsnLysLysArgThrAlaLeuIleLysAlaValGlnCysGl 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     601 GTCCTTGACAACAAAAAGAGGACAGCTCTGATAAAGGCCGTACAATGCCA 650
```

```
     217 nGluAspGluCysAlaLeuMetLeuLeuGluHisGlyThrAspProAsnI 234
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     651 GGAAGATGAATGTGCGTTAATGTTGCTGGAACATGGCACTGATCCAAATA 700

     234 leProAspGluTyrGlyAsnThrThrLeuHisTyrAlaIleTyrAsnGlu 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     701 TTCCAGATGAGTATGGAAATACCACTCTGCACTACGCTATCTATAATGAA 750

     251 AspLysLeuMetAlaLysAlaLeuLeuLeuTyrGlyAlaAspIleGluSe 267
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     751 GATAAATTAATGGCCAAAGCACTGCTCTTATATGGTGCTGATATCGAATC 800

     267 rLysAsnLysHisGlyLeuThrProLeuLeuLeuGlyValHisGluGlnL 284
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     801 AAAAAACAAGCATGGCCTCACACCACTGTTACTTGGTGTACATGAGCAAA 850

     284 ysGlnGlnValValLysPheLeuIleLysLysLysAlaAsnLeuAsnAla 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     851 AACAGCAAGTCGTGAAATTTTTAATCAAGAAAAAAGCGAATTTAAATGCA 900

     301 LeuAspArgTyrGlyArgThrAlaLeuIleLeuAlaValCysCysGlySe 317
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     901 CTGGATAGATATGGAAGGACTGCTCTCATACTTGCTGTATGTTGTGGATC 950

     317 rAlaSerIleValSerLeuLeuLeuGluGlnAsnIleAspValSerSerG 334
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     951 AGCAAGTATAGTCAGCCTTCTACTTGAGCAAAATATTGATGTATCTTCTC 1000

     334 lnAspLeuSerGlyGlnThrAlaArgGluTyrAlaValSerSerHisHis 350
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    1001 AAGATCTATCTGGACAGACGGCCAGAGAGTATGCTGTTTCTAGTCATCAT 1050

     351 HisValIleCysGlnLeuLeuSerAspTyrLysGluLysGlnMetLeuLy 367
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    1051 CATGTAATTTGCCAGTTACTTTCTGACTACAAAGAAAAACAGATGCTAAA 1100

     367 sIleSerSerGluAsnSerAsnProGluAsnValSerArgThrArgAsnL 384
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
    1101 AATCTCTTCTGAAAACAGCAATCCAGAAAATGTCTCAAGAACCAGAAATA 1150

     384 ys 384
         ||
    1151 AA 1152

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAH02779

seq_documentation_block:
ID   AAH02779 standard; cDNA; 1155 BP.
XX
AC   AAH02779;
XX
DT   14-JUN-2001  (first entry)
XX
DE   Prostate tumour antigen determined cDNA splice variant of B305D #8.
XX
KW   Human; prostate tumour antigen; prostate tumour; therapy; diagnosis;
KW   prostate cancer; immunogenic; cytostatic; vaccine; ss.
XX
OS   Homo sapiens.
XX
PN   WO200125272-A2.
XX
PD   12-APR-2001.
XX
PF   04-OCT-2000; 2000WO-US27464.
XX
PR   04-OCT-1999;   99US-0157455.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Xu J,  Skeiky YAW,  Reed SG,  Cheever MA;
XX
DR   WPI; 2001-245062/25.
DR   P-PSDB; AAB74815.
XX
PT   Prostate specific protein and its encoding polynucleotide, useful for
PT   the treatment and diagnosis of prostate cancer -
XX
PS   Claim 50; Page 231-232; 276pp; English.
XX
CC   The present invention describes an isolated polypeptide (I) comprising
CC   at least an immunogenic portion of a prostate tumour antigen protein or
CC   its variant. (I) have cytostatic activity and can be used in vaccine
CC   production. (I), prostate tumour antigen polynucleotides, an antigen
CC   presenting cell (APC e.g. a dendritic cell) that expresses (I), and a
CC   pharmaceutical composition containing (I) are useful for inhibiting the
CC   development of cancer in a patient. Antibodies specific for prostate
CC   specific proteins and oligonucleotides that hybridise to a
CC   polynucleotide that encodes a prostate specific protein are useful
CC   for detecting the presence or absence of a cancer or monitoring the
CC   progression the progression of a cancer, especially prostate cancer.
CC   AAH02422 to AAH2872, AAB74798 to AAB74821 and AAB74830 are sequences
CC   used in the exemplification of the present invention.
XX
SQ   Sequence 1155 BP; 346 A; 253 C; 297 G; 259 T; 0 other;


alignment_scores:
           Quality: 2064.00              Length:    384
             Ratio:  5.375                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-810-936-304 x AAH02779   ..

 Align seg 1/1  to: AAH02779  from: 1  to: 1155

       1 MetValValGluValAspSerMetProAlaAlaSerSerValLysLysPr 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
       1 ATGGTGGTTGAGGTTGATTCCATGCCGGCTGCCTCTTCTGTGAAGAAGCC 50

      17 oPheGlyLeuArgSerLysMetGlyLysTrpCysCysArgCysPheProC 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
      51 ATTTGGTCTCAGGAGCAAGATGGGCAAGTGGTGCTGCCGTTGCTTCCCCT 100

      34 ysCysArgGluSerGlyLysSerAsnValGlyThrSerGlyAspHisAsp 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     101 GCTGCAGGGAGAGCGGCAAGAGCAACGTGGGCACTTCTGGAGACCACGAC 150

      51 AspSerAlaMetLysThrLeuArgSerLysMetGlyLysTrpCysArgHi 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     151 GACTCTGCTATGAAGACACTCAGGAGCAAGATGGGCAAGTGGTGCCGCCA 200

      67 sCysPheProCysCysArgGlySerGlyLysSerAsnValGlyAlaSerG 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     201 CTGCTTCCCCTGCTGCAGGGGGAGTGGCAAGAGCAACGTGGGCGCTTCTG 250

      84 lyAspHisAspAspSerAlaMetLysThrLeuArgAsnLysMetGlyLys 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     251 GAGACCACGACGACTCTGCTATGAAGACACTCAGGAACAAGATGGGCAAG 300

     101 TrpCysCysHisCysPheProCysCysArgGlySerGlyLysSerLysVa 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     301 TGGTGCTGCCACTGCTTCCCCTGCTGCAGGGGGAGCGGCAAGAGCAAGGT 350

     117 lGlyAlaTrpGlyAspTyrAspAspSerAlaPheMetGluProArgTyrH 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     351 GGGCGCTTGGGGAGACTACGATGACAGTGCCTTCATGGAGCCCAGGTACC 400

     134 isValArgGlyGluAspLeuAspLysLeuHisArgAlaAlaTrpTrpGly 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     401 ACGTCCGTGGAGAAGATCTGGACAAGCTCCACAGAGCTGCCTGGTGGGGT 450

     151 LysValProArgLysAspLeuIleValMetLeuArgAspThrAspValAs 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
     451 AAAGTCCCCAGAAAGGATCTCATCGTCATGCTCAGGGACACTGACGTGAA 500

     167 nLysLysAspLysGlnLysArgThrAlaLeuHisLeuAlaSerAlaAsnG 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     501 CAAGAAGGACAAGCAAAAGAGGACTGCTCTACATCTGGCCTCTGCCAATG 550

     184 lyAsnSerGluValValLysLeuLeuLeuAspArgArgCysGlnLeuAsn 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     551 GGAATTCAGAAGTAGTAAAACTCCTGCTGGACAGACGATGTCAACTTAAT 600

     201 ValLeuAspAsnLysLysArgThrAlaLeuIleLysAlaValGlnCysGl 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     601 GTCCTTGACAACAAAAAGAGGACAGCTCTGATAAAGGCCGTACAATGCCA 650

     217 nGluAspGluCysAlaLeuMetLeuLeuGluHisGlyThrAspProAsnI 234
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     651 GGAAGATGAATGTGCGTTAATGTTGCTGGAACATGGCACTGATCCAAATA 700

     234 leProAspGluTyrGlyAsnThrThrLeuHisTyrAlaIleTyrAsnGlu 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     701 TTCCAGATGAGTATGGAAATACCACTCTGCACTACGCTATCTATAATGAA 750

     251 AspLysLeuMetAlaLysAlaLeuLeuLeuTyrGlyAlaAspIleGluSe 267
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     751 GATAAATTAATGGCCAAAGCACTGCTCTTATATGGTGCTGATATCGAATC 800

     267 rLysAsnLysHisGlyLeuThrProLeuLeuLeuGlyValHisGluGlnL 284
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     801 AAAAAACAAGCATGGCCTCACACCACTGTTACTTGGTGTACATGAGCAAA 850

     284 ysGlnGlnValValLysPheLeuIleLysLysLysAlaAsnLeuAsnAla 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     851 AACAGCAAGTCGTGAAATTTTTAATCAAGAAAAAAGCGAATTTAAATGCA 900

     301 LeuAspArgTyrGlyArgThrAlaLeuIleLeuAlaValCysCysGlySe 317
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     901 CTGGATAGATATGGAAGGACTGCTCTCATACTTGCTGTATGTTGTGGATC 950

     317 rAlaSerIleValSerLeuLeuLeuGluGlnAsnIleAspValSerSerG 334
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     951 AGCAAGTATAGTCAGCCTTCTACTTGAGCAAAATATTGATGTATCTTCTC 1000

     334 lnAspLeuSerGlyGlnThrAlaArgGluTyrAlaValSerSerHisHis 350
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1001 AAGATCTATCTGGACAGACGGCCAGAGAGTATGCTGTTTCTAGTCATCAT 1050

     351 HisValIleCysGlnLeuLeuSerAspTyrLysGluLysGlnMetLeuLy 367
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1051 CATGTAATTTGCCAGTTACTTTCTGACTACAAAGAAAAACAGATGCTAAA 1100

     367 sIleSerSerGluAsnSerAsnProGluAsnValSerArgThrArgAsnL 384
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1101 AATCTCTTCTGAAAACAGCAATCCAGAAAATGTCTCAAGAACCAGAAATA 1150

     384 ys 384
         ||
    1151 AA 1152

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2002.DAT:AAS99857

seq_documentation_block:
ID   AAS99857 standard; cDNA; 1155 BP.
XX
AC   AAS99857;
XX
DT   12-MAR-2002  (first entry)
XX
DE   Breast tumour-specific DNA B11Ag1 splice variant B11C-15.
XX
KW   Human; breast cancer; PCR primer; ss; cytostatic; immunostimulant;
KW   tumour; vaccine; immunogenic.
XX
OS   Homo sapiens.
XX
PN   WO200190152-A2.
XX
PD   29-NOV-2001.
XX
PF   22-MAY-2001; 2001WO-US16776.
XX
PR   24-MAY-2000; 2000US-0577505.
PR   08-JUN-2000; 2000US-0590583.
PR   26-OCT-2000; 2000US-0699295.
PR   16-MAR-2001; 2001US-0810936.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Frudakis TN,  Reed SG,  Smith JM,  Misher LE,  Dillon DC,  Retter MW;
PI   Wang A,  Skeiky YAW,  Harlocker SL,  Day CH;
XX
DR   WPI; 2002-089919/12.
DR   P-PSDB; AAU74377.
XX
PT   New breast tumour proteins and polynucleotides encoding them, useful for
PT   treating and/or preventing cancer, particularly breast cancer, and for
PT   eliciting humoral and/or cellular immune response  -
XX
PS   Claim 1; Page 223; 245pp; English.
XX
CC   The invention relates to novel breast tumour polynucleotides and
CC   polypeptides. The polypeptides and polynucleotides are useful in
CC   pharmaceutical compositions for treating and/or preventing cancer,
CC   particularly breast cancer, and for eliciting an immune response,
CC   particularly humoral and/or cellular immune response. The polynucleotides
CC   may be used as probes or primers for nucleic acid hybridisation, in the
CC   design and preparation of ribozyme molecules for inhibiting expression of
CC   tumour polypeptides and proteins, and in recombinant DNA molecules to
CC   direct expression of a polypeptide in host cells. AAS99570-AAS99888
CC   represent novel human breast cancer protein coding sequences and
CC   PCR primers of the invention.
XX
SQ   Sequence 1155 BP; 346 A; 253 C; 297 G; 259 T; 0 other;


alignment_scores:
            Quality: 2064.00              Length:    384
              Ratio:  5.375                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-810-936-304 x AAS99857   ..

 Align seg 1/1  to: AAS99857  from: 1  to: 1155

       1 MetValValGluValAspSerMetProAlaAlaSerSerValLysLysPr 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 ATGGTGGTTGAGGTTGATTCCATGCCGGCTGCCTCTTCTGTGAAGAAGCC 50

      17 oPheGlyLeuArgSerLysMetGlyLysTrpCysCysArgCysPheProC 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 ATTTGGTCTCAGGAGCAAGATGGGCAAGTGGTGCTGCCGTTGCTTCCCCT 100

      34 ysCysArgGluSerGlyLysSerAsnValGlyThrSerGlyAspHisAsp 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 GCTGCAGGGAGAGCGGCAAGAGCAACGTGGGCACTTCTGGAGACCACGAC 150

      51 AspSerAlaMetLysThrLeuArgSerLysMetGlyLysTrpCysArgHi 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     151 GACTCTGCTATGAAGACACTCAGGAGCAAGATGGGCAAGTGGTGCCGCCA 200

      67 sCysPheProCysCysArgGlySerGlyLysSerAsnValGlyAlaSerG 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     201 CTGCTTCCCCTGCTGCAGGGGGAGTGGCAAGAGCAACGTGGGCGCTTCTG 250

      84 lyAspHisAspAspSerAlaMetLysThrLeuArgAsnLysMetGlyLys 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
 251 GAGACCACGACGACTCTGCTATGAAGACACTCAGGAACAAGATGGGCAAG 300

 101 TrpCysCysHisCysPheProCysCysArgGlySerGlyLysSerLysVa 117
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 301 TGGTGCTGCCACTGCTTCCCCTGCTGCAGGGGGAGCGGCAAGAGCAAGGT 350

 117 lGlyAlaTrpGlyAspTyrAspAspSerAlaPheMetGluProArgTyrH 134
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 351 GGGCGCTTGGGGAGACTACGATGACAGTGCCTTCATGGAGCCCAGGTACC 400

 134 isValArgGlyGluAspLeuAspLysLeuHisArgAlaAlaTrpTrpGly 150
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 401 ACGTCCGTGGAGAAGATCTGGACAAGCTCCACAGAGCTGCCTGGTGGGGT 450

 151 LysValProArgLysAspLeuIleValMetLeuArgAspThrAspValAs 167
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 451 AAAGTCCCCAGAAAGGATCTCATCGTCATGCTCAGGGACACTGACGTGAA 500

 167 nLysLysAspLysGlnLysArgThrAlaLeuHisLeuAlaSerAlaAsnG 184
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 501 CAAGAAGGACAAGCAAAAGAGGACTGCTCTACATCTGGCCTCTGCCAATG 550

 184 lyAsnSerGluValValLysLeuLeuLeuAspArgArgCysGlnLeuAsn 200
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 551 GGAATTCAGAAGTAGTAAAACTCCTGCTGGACAGACGATGTCAACTTAAT 600

 201 ValLeuAspAsnLysLysArgThrAlaLeuIleLysAlaValGlnCysGl 217
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 601 GTCCTTGACAACAAAAAGAGGACAGCTCTGATAAAGGCCGTACAATGCCA 650

 217 nGluAspGluCysAlaLeuMetLeuLeuGluHisGlyThrAspProAsnI 234
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 651 GGAAGATGAATGTGCGTTAATGTTGCTGGAACATGGCACTGATCCAAATA 700

 234 leProAspGluTyrGlyAsnThrThrLeuHisTyrAlaIleTyrAsnGlu 250
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 701 TTCCAGATGAGTATGGAAATACCACTCTGCACTACGCTATCTATAATGAA 750

 251 AspLysLeuMetAlaLysAlaLeuLeuLeuTyrGlyAlaAspIleGluSe 267
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 751 GATAAATTAATGGCCAAAGCACTGCTCTTATATGGTGCTGATATCGAATC 800

 267 rLysAsnLysHisGlyLeuThrProLeuLeuLeuGlyValHisGluGlnL 284
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 801 AAAAAACAAGCATGGCCTCACACCACTGTTACTTGGTGTACATGAGCAAA 850

 284 ysGlnGlnValValLysPheLeuIleLysLysLysAlaAsnLeuAsnAla 300
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 851 AACAGCAAGTCGTGAAATTTTTAATCAAGAAAAAAGCGAATTTAAATGCA 900

 301 LeuAspArgTyrGlyArgThrAlaLeuIleLeuAlaValCysCysGlySe 317
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 901 CTGGATAGATATGGAAGGACTGCTCTCATACTTGCTGTATGTTGTGGATC 950

 317 rAlaSerIleValSerLeuLeuLeuGluGlnAsnIleAspValSerSerG 334
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 951 AGCAAGTATAGTCAGCCTTCTACTTGAGCAAAATATTGATGTATCTTCTC 1000

 334 lnAspLeuSerGlyGlnThrAlaArgGluTyrAlaValSerSerHisHis 350
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1001 AAGATCTATCTGGACAGACGGCCAGAGAGTATGCTGTTTCTAGTCATCAT 1050

 351 HisValIleCysGlnLeuLeuSerAspTyrLysGluLysGlnMetLeuLy 367
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1051 CATGTAATTTGCCAGTTACTTTCTGACTACAAAGAAAAACAGATGCTAAA 1100

 367 sIleSerSerGluAsnSerAsnProGluAsnValSerArgThrArgAsnL 384
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1101 AATCTCTTCTGAAAACAGCAATCCAGAAAATGTCTCAAGAACCAGAAATA 1150

 384 ys 384
     ||
1151 AA 1152
```

```
seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2002.DAT:AAS99869

seq_documentation_block:
ID   AAS99869 standard; DNA; 1590 BP.
XX
AC   AAS99869;
XX
DT   12-MAR-2002  (first entry)
XX
DE   Breast tumour-specific gene B305D fusion construct.
XX
KW   Human; breast cancer; PCR primer; ss; cytostatic; immunostimulant;
KW   tumour; vaccine; immunogenic.
XX
OS   Homo sapiens.
XX
PN   WO200190152-A2.
XX
PD   29-NOV-2001.
XX
PF   22-MAY-2001; 2001WO-US16776.
XX
PR   24-MAY-2000; 2000US-0577505.
PR   08-JUN-2000; 2000US-0590583.
PR   26-OCT-2000; 2000US-0699295.
PR   16-MAR-2001; 2001US-0810936.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Frudakis TN,  Reed SG,  Smith JM,  Misher LE,  Dillon DC,  Retter MW;
PI   Wang A,  Skeiky YAW,  Harlocker SL,  Day CH;
XX
DR   WPI; 2002-089919/12.
XX
PT   New breast tumour proteins and polynucleotides encoding them, useful for
PT   treating and/or preventing cancer, particularly breast cancer, and for
PT   eliciting humoral and/or cellular immune response  -
XX
PS   Example 8; Page 235; 245pp; English.
XX
CC   The invention relates to novel breast tumour polynucleotides and
CC   polypeptides. The polypeptides and polynucleotides are useful in
CC   pharmaceutical compositions for treating and/or preventing cancer,
CC   particularly breast cancer, and for eliciting an immune response,
CC   particularly humoral and/or cellular immune response. The polynucleotides
CC   may be used as probes or primers for nucleic acid hybridisation, in the
CC   design and preparation of ribozyme molecules for inhibiting expression of
CC   tumour polypeptides and proteins, and in recombinant DNA molecules to
CC   direct expression of a polypeptide in host cells. AAS99570-AAS99888
CC   represent novel human breast cancer protein coding sequences and
CC   PCR primers of the invention.
XX
SQ   Sequence 1590 BP; 424 A; 403 C; 433 G; 329 T; 1 other;


alignment_scores:
            Quality: 2064.00              Length:    384
              Ratio:  5.375                Gaps:      0
 Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
 US-09-810-936-304 x AAS99869   ..

 Align seg 1/1  to: AAS99869  from: 1  to: 1590

       1 MetValValGluValAspSerMetProAlaAlaSerSerValLysLysPr 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     436 ATGGTGGTTGAGGTTGATTCCATGCCGGCTGCTTCTTCTGTGAAGAAGCC 485

      17 oPheGlyLeuArgSerLysMetGlyLysTrpCysCysArgCysPheProC 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     486 ATTTGGTCTCAGGAGCAAGATGGGCAAGTGGTGCTGCCGTTGCTTCCCCT 535
```

```
 34 ysCysArgGluSerGlyLysSerAsnValGlyThrSerGlyAspHisAsp 50
    ||||||||||||||||||||||||||||||||||||||||||||||||||
536 GCTGCAGGGAGAGCGGCAAGAGCAACGTGGGCACTTCTGGAGACCACGAC 585

 51 AspSerAlaMetLysThrLeuArgSerLysMetGlyLysTrpCysArgHi 67
    ||||||||||||||||||||||||||||||||||||||||||||||||||
586 GACTCTGCTATGAAGACACTCAGGAGCAAGATGGGCAAGTGGTGCCGCCA 635

 67 sCysPheProCysCysArgGlySerGlyLysSerAsnValGlyAlaSerG 84
    ||||||||||||||||||||||||||||||||||||||||||||||||||
636 CTGCTTCCCCTGCTGCAGGGGGAGTGGCAAGAGCAACGTGGGCGCTTCTG 685

 84 lyAspHisAspAspSerAlaMetLysThrLeuArgAsnLysMetGlyLys 100
    ||||||||||||||||||||||||||||||||||||||||||||||||||
686 GAGACCACGACGACTCTGCTATGAAGACACTCAGGAACAAGATGGGCAAG 735

101 TrpCysCysHisCysPheProCysCysArgGlySerGlyLysSerLysVa 117
    ||||||||||||||||||||||||||||||||||||||||||||||||||
736 TGGTGCTGCCACTGCTTCCCCTGCTGCAGGGGGAGCGGCAAGAGCAAGGT 785

117 lGlyAlaTrpGlyAspTyrAspAspSerAlaPheMetGluProArgTyrH 134
    ||||||||||||||||||||||||||||||||||||||||||||||||||
786 GGGCGCTTGGGGAGACTACGATGACAGYGCCTTCATGGAGCCCAGGTACC 835

134 isValArgGlyGluAspLeuAspLysLeuHisArgAlaAlaTrpTrpGly 150
    ||||||||||||||||||||||||||||||||||||||||||||||||||
836 ACGTCCGTGGAGAAGATCTGGACAAGCTCCACAGAGCTGCCTGGTGGGGT 885

151 LysValProArgLysAspLeuIleValMetLeuArgAspThrAspValAs 167
    ||||||||||||||||||||||||||||||||||||||||||||||||||
886 AAAGTCCCCAGAAAGGATCTCATCGTCATGCTCAGGGACACTGACGTGAA 935

167 nLysLysAspLysGlnLysArgThrAlaLeuHisLeuAlaSerAlaAsnG 184
    ||||||||||||||||||||||||||||||||||||||||||||||||||
936 CAAGAAGGACAAGCAAAAGAGGACTGCTCTACATCTGGCCTCTGCCAATG 985

184 lyAsnSerGluValValLysLeuLeuLeuAspArgArgCysGlnLeuAsn 200
    ||||||||||||||||||||||||||||||||||||||||||||||||||
986 GGAATTCAGAAGTAGTAAAACTCCTGCTGGACAGACGATGTCAACTTAAT 1035

201 ValLeuAspAsnLysLysArgThrAlaLeuIleLysAlaValGlnCysGl 217
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1036 GTCCTTGACAACAAAAAGAGGACAGCTCTGATAAAGGCCGTACAATGCCA 1085

217 nGluAspGluCysAlaLeuMetLeuLeuGluHisGlyThrAspProAsnI 234
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1086 GGAAGATGAATGTGCGTTAATGTTGCTGGAACATGGCACTGATCCAAATA 1135

234 leProAspGluTyrGlyAsnThrThrLeuHisTyrAlaIleTyrAsnGlu 250
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1136 TTCCAGATGAGTATGGAAATACCACTCTGCACTACGCTATCTATAATGAA 1185

251 AspLysLeuMetAlaLysAlaLeuLeuLeuTyrGlyAlaAspIleGluSe 267
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1186 GATAAATTAATGGCCAAAGCACTGCTCTTATATGGTGCTGATATCGAATC 1235

267 rLysAsnLysHisGlyLeuThrProLeuLeuLeuGlyValHisGluGlnL 284
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1236 AAAAAACAAGCATGGCCTCACACCACTGTTACTTGGTGTACATGAGCAAA 1285

284 ysGlnGlnValValLysPheLeuIleLysLysLysAlaAsnLeuAsnAla 300
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1286 AACAGCAAGTCGTGAAATTTTTAATCAAGAAAAAAGCGAATTTAAATGCA 1335

301 LeuAspArgTyrGlyArgThrAlaLeuIleLeuAlaValCysCysGlySe 317
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1336 CTGGATAGATATGGAAGGACTGCTCTCATACTTGCTGTATGTTGTGGATC 1385

317 rAlaSerIleValSerLeuLeuLeuGluGlnAsnIleAspValSerSerG 334
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1386 AGCAAGTATAGTCAGCCTTCTACTTGAGCAAAATATTGATGTATCTTCTC 1435

334 lnAspLeuSerGlyGlnThrAlaArgGluTyrAlaValSerSerHisHis 350
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1436 AAGATCTATCTGGACAGACGGCCAGAGAGTATGCTGTTTCTAGTCATCAT 1485

351 HisValIleCysGlnLeuLeuSerAspTyrLysGluLysGlnMetLeuLy 367
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1486 CATGTAATTTGCCAGTTACTTTCTGACTACAAAGAAAAACAGATGCTAAA 1535

367 sIleSerSerGluAsnSerAsnProGluAsnValSerArgThrArgAsnL 384
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1536 AATCTCTTCTGAAAACAGCAATCCAGAAAATGTCTCAAGAACCAGAAATA 1585

384 ys 384
    ||
1586 AA 1587
```

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2002.DAT:AAS99872

seq_documentation_block:

ID AAS99872 standard; DNA; 1155 BP.
XX
AC AAS99872;
XX
DT 12-MAR-2002 (first entry)
XX
DE Breast tumour-specific gene B305D homologue #2.
XX
KW Human; breast cancer; PCR primer; ss; cytostatic; immunostimulant;
KW tumour; vaccine; immunogenic.
XX
OS Homo sapiens.
XX
PN WO200190152-A2.
XX
PD 29-NOV-2001.
XX
PF 22-MAY-2001; 2001WO-US16776.
XX
PR 24-MAY-2000; 2000US-0577505.
PR 08-JUN-2000; 2000US-0590583.
PR 26-OCT-2000; 2000US-0699295.
PR 16-MAR-2001; 2001US-0810936.
XX
PA (CORI-) CORIXA CORP.
XX
PI Frudakis TN, Reed SG, Smith JM, Misher LE, Dillon DC, Retter MW;
PI Wang A, Skeiky YAW, Harlocker SL, Day CH;
XX
DR WPI; 2002-089919/12.
DR P-PSDB; AAU74390.
XX
PT New breast tumour proteins and polynucleotides encoding them, useful for
PT treating and/or preventing cancer, particularly breast cancer, and for
PT eliciting humoral and/or cellular immune response -
XX
PS Claim 1; Page 239; 245pp; English.
XX
CC The invention relates to novel breast tumour polynucleotides and
CC polypeptides. The polypeptides and polynucleotides are useful in
CC pharmaceutical compositions for treating and/or preventing cancer,
CC particularly breast cancer, and for eliciting an immune response,
CC particularly humoral and/or cellular immune response. The polynucleotides
CC may be used as probes or primers for nucleic acid hybridisation, in the
CC design and preparation of ribozyme molecules for inhibiting expression of
CC tumour polypeptides and proteins, and in recombinant DNA molecules to
CC direct expression of a polypeptide in host cells. AAS99570-AAS99888
CC represent novel human breast cancer protein coding sequences and
CC PCR primers of the invention.
XX
SQ Sequence 1155 BP; 346 A; 253 C; 296 G; 260 T; 0 other;

alignment_scores:

```
          Quality: 2054.00              Length:    384
            Ratio:  5.349                Gaps:      0
Percent Similarity: 100.000   Percent Identity: 99.479

alignment_block:
US-09-810-936-304 x AAS99872   ..

Align seg 1/1  to: AAS99872  from: 1  to: 1155

       1 MetValValGluValAspSerMetProAlaAlaSerSerValLysLysPr 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 ATGGTGGTTGAGGTTGATTCCATGCCGGCTGCCTCTTCTGTGAAGAAGCC 50

      17 oPheGlyLeuArgSerLysMetGlyLysTrpCysCysArgCysPheProC 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 ATTTGGTCTCAGGAGCAAGATGGGCAAGTGGTGCTGCCGTTGCTTCCCCT 100

      34 ysCysArgGluSerGlyLysSerAsnValGlyThrSerGlyAspHisAsp 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 GCTGCAGGGAGAGCGGCAAGAGCAACGTGGGCACTTCTGGAGACCACGAC 150

      51 AspSerAlaMetLysThrLeuArgSerLysMetGlyLysTrpCysArgHi 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     151 GACTCTGCTATGAAGACACTCAGGAGCAAGATGGGCAAGTGGTGCCGCCA 200

      67 sCysProProCysCysArgGlySerGlyLysSerAsnValGlyAlaSerG 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     201 CTGCTTCCCCTGCTGCAGGGGGAGTGGCAAGAGCAACGTGGGCGCTTCTG 250

      84 lyAspHisAspAspSerAlaMetLysThrLeuArgAsnLysMetGlyLys 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     251 GAGACCACGACGACTCTGCTATGAAGACACTCAGGAACAAGATGGGCAAG 300

     101 TrpCysCysHisCysProProCysCysArgGlySerGlyLysSerLysVa 117
         ||||||||||||||||||||||||||||||||||||||:::||||||||||
     301 TGGTGCTGCCACTGCTTCCCCTGCTGCAGGGGGAGCAGCAAGAGCAAGGT 350

     117 lGlyAlaTrpGlyAspTyrAspAspSerAlaPheMetGluProArgTyrH 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     351 GGGCGCTTGGGGAGACTACGATGACAGTGCCTTCATGGAGCCCAGGTACC 400

     134 isValArgGlyGluAspLeuAspLysLeuHisArgAlaAlaTrpTrpGly 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     401 ACGTCCGTGGAGAAGATCTGGACAAGCTCCACAGAGCTGCCTGGTGGGGT 450

     151 LysValProArgLysAspLeuIleValMetLeuArgAspThrAspValAs 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     451 AAAGTCCCCAGAAAGGATCTCATCGTCATGCTCAGGGACACTGACGTGAA 500

     167 nLysLysAspLysGlnLysArgThrAlaLeuHisLeuAlaSerAlaAsnG 184
         ||||:::||||||||||||||||||||||||||||||||||||||||||||
     501 CAAGCAGGACAAGCAAAAGAGGACTGCTCTACATCTGGCCTCTGCCAATG 550

     184 lyAsnSerGluValValLysLeuLeuLeuAspArgArgCysGlnLeuAsn 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     551 GGAATTCAGAAGTAGTAAAACTCCTGCTGGACAGACGATGTCAACTTAAT 600

     201 ValLeuAspAsnLysLysArgThrAlaLeuIleLysAlaValGlnCysGl 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     601 GTCCTTGACAACAAAAAGAGGACAGCTCTGATAAAGGCCGTACAATGCCA 650

     217 nGluAspGluCysAlaLeuMetLeuLeuGluHisGlyThrAspProAsnI 234
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     651 GGAAGATGAATGTGCGTTAATGTTGCTGGAACATGGCACTGATCCAAATA 700

     234 leProAspGluTyrGlyAsnThrThrLeuHisTyrAlaIleTyrAsnGlu 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     701 TTCCAGATGAGTATGGAAATACCACTCTGCACTACGCTATCTATAATGAA 750

     251 AspLysLeuMetAlaLysAlaLeuLeuLeuTyrGlyAlaAspIleGluSe 267
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     751 GATAAATTAATGGCCAAAGCACTGCTCTTATATGGTGCTGATATCGAATC 800

     267 rLysAsnLysHisGlyLeuThrProLeuLeuLeuGlyValHisGluGlnL 284
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     801 AAAAAACAAGCATGGCCTCACACCACTGTTACTTGGTGTACATGAGCAAA 850

     284 ysGlnGlnValValLysPheLeuIleLysLysLysAlaAsnLeuAsnAla 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     851 AACAGCAAGTCGTGAAATTTTTAATTAAGAAAAAAGCGAATTTAAATGCA 900

     301 LeuAspArgTyrGlyArgThrAlaLeuIleLeuAlaValCysCysGlySe 317
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     901 CTGGATAGATATGGAAGGACTGCTCTCATACTTGCTGTATGTTGTGGATC 950

     317 rAlaSerIleValSerLeuLeuLeuGluGlnAsnIleAspValSerSerG 334
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     951 AGCAAGTATAGTCAGCCTTCTACTTGAGCAAAATATTGATGTATCTTCTC 1000

     334 lnAspLeuSerGlyGlnThrAlaArgGluTyrAlaValSerSerHisHis 350
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1001 AAGATCTATCTGGACAGACGGCCAGAGAGTATGCTGTTTCTAGTCATCAT 1050

     351 HisValIleCysGlnLeuLeuSerAspTyrLysGluLysGlnMetLeuLy 367
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1051 CATGTAATTTGCCAGTTACTTTCTGACTACAAAGAAAAACAGATGCTAAA 1100

     367 sIleSerSerGluAsnSerAsnProGluAsnValSerArgThrArgAsnL 384
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1101 AATCTCTTCTGAAAACAGCAATCCAGAAAATGTCTCAAGAACCAGAAATA 1150

     384 ys 384
         ||
    1151 AA 1152

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:AAC81012

seq_documentation_block:
ID   AAC81012 standard; cDNA; 2000 BP.
XX
AC   AAC81012;
XX
DT   13-FEB-2001  (first entry)
XX
DE   Human B11Ag1 antigen splice isoform B11C-8 cDNA.
XX
KW   Human; breast tumour-specific antigen; cytostatic; vaccine;
KW   breast cancer; B18Ag1; B11Ag1; B15Ag1; ss.
XX
OS   Homo sapiens.
XX
PN   WO200061753-A2.
XX
PD   19-OCT-2000.
XX
PF   07-APR-2000; 2000WO-US09312.
XX
PR   09-APR-1999;   99US-0289198.
PR   28-OCT-1999;   99US-0429755.
PR   23-MAR-2000; 2000US-0534825.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Frudakis TN,  Smith JM,  Reed SG,  Misher LE,  Retter MW,  Dillon DC;
XX
DR   WPI; 2000-628403/60.
DR   P-PSDB; AAB28629.
XX
PT   An isolated polypeptide comprising an immunogenic portion of a breast
PT   tumor protein used for inhibiting the development of cancer, especially
PT   breast cancer, and monitoring cancer progression in a patient -
XX
PS   Claim 4; Page 177-178; 187pp; English.
XX
CC   The present sequence is given in a specification relating to compositions
CC   and methods for the treatment and diagnosis of breast cancer. Nucleotide
```

```
CC   sequences that are preferentially expressed in breast tumour tissue, and
CC   the polypeptides encoded by such nucleotide sequences, are used in
CC   compositions and vaccines to inhibit the development of cancer,
CC   especially breast cancer. The progression of a cancer may be monitored by
CC   carrying out detection of tumour-specific antigens at subsequent time
CC   points and comparing the results from the different time points.
CC   CD4+ and/or CD8+ T-Cells isolated from the cancer patient may be treated
CC   with tumour-specific polypeptides, polynucleotides encoding the
CC   polypeptides or antigen presenting cells expressing the polypeptides. The
CC   cells are then administered to the patient to inhibit development of
CC   cancer.
XX
SQ   Sequence 2000 BP; 698 A; 388 C; 489 G; 425 T; 0 other;


alignment_scores:
            Quality: 2024.00              Length:    376
              Ratio:  5.383                Gaps:      0
 Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
 US-09-810-936-304 x AAC81012   ..

 Align seg 1/1  to: AAC81012  from: 1  to: 2000

       1 MetValValGluValAspSerMetProAlaAlaSerSerValLysLysPr 17
         |||||||||||||||||||||||||||||||||||||||||||||||||||
       1 ATGGTGGTTGAGGTTGATTCCATGCCGGCTGCCTCTTCTGTGAAGAAGCC 50

      17 oPheGlyLeuArgSerLysMetGlyLysTrpCysCysArgCysPheProC 34
         |||||||||||||||||||||||||||||||||||||||||||||||||||
      51 ATTTGGTCTCAGGAGCAAGATGGGCAAGTGGTGCTGCCGTTGCTTCCCCT 100

      34 ysCysArgGluSerGlyLysSerAsnValGlyThrSerGlyAspHisAsp 50
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     101 GCTGCAGGGAGAGCGGCAAGAGCAACGTGGGCACTTCTGGAGACCACGAC 150

      51 AspSerAlaMetLysThrLeuArgSerLysMetGlyLysTrpCysArgHi 67
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     151 GACTCTGCTATGAAGACACTCAGGAGCAAGATGGGCAAGTGGTGCCGCCA 200

      67 sCysPheProCysCysArgGlySerGlyLysSerAsnValGlyAlaSerG 84
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     201 CTGCTTCCCCTGCTGCAGGGGGAGTGGCAAGAGCAACGTGGGCGCTTCTG 250

      84 lyAspHisAspAspSerAlaMetLysThrLeuArgAsnLysMetGlyLys 100
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     251 GAGACCACGACGACTCTGCTATGAAGACACTCAGGAACAAGATGGGCAAG 300

     101 TrpCysCysHisCysPheProCysCysArgGlySerGlyLysSerLysVa 117
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     301 TGGTGCTGCCACTGCTTCCCCTGCTGCAGGGGGAGCGGCAAGAGCAAGGT 350

     117 lGlyAlaTrpGlyAspTyrAspAspSerAlaPheMetGluProArgTyrH 134
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     351 GGGCGCTTGGGGAGACTACGATGACAGTGCCTTCATGGAGCCCAGGTACC 400

     134 isValArgGlyGluAspLeuAspLysLeuHisArgAlaAlaTrpTrpGly 150
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     401 ACGTCCGTGGAGAAGATCTGGACAAGCTCCACAGAGCTGCCTGGTGGGGT 450

     151 LysValProArgLysAspLeuIleValMetLeuArgAspThrAspValAs 167
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     451 AAAGTCCCCAGAAAGGATCTCATCGTCATGCTCAGGGACACTGACGTGAA 500

     167 nLysLysAspLysGlnLysArgThrAlaLeuHisLeuAlaSerAlaAsnG 184
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     501 CAAGAAGGACAAGCAAAAGAGGACTGCTCTACATCTGGCCTCTGCCAATG 550

     184 lyAsnSerGluValValLysLeuLeuLeuAspArgArgCysGlnLeuAsn 200
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     551 GGAATTCAGAAGTAGTAAAACTCCTGCTGGACAGACGATGTCAACTTAAT 600

     201 ValLeuAspAsnLysLysArgThrAlaLeuIleLysAlaValGlnCysGl 217
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     601 GTCCTTGACAACAAAAAGAGGACAGCTCTGATAAAGGCCGTACAATGCCA 650

     217 nGluAspGluCysAlaLeuMetLeuLeuGluHisGlyThrAspProAsnI 234
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     651 GGAAGATGAATGTGCGTTAATGTTGCTGGAACATGGCACTGATCCAAATA 700

     234 leProAspGluTyrGlyAsnThrThrLeuHisTyrAlaIleTyrAsnGlu 250
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     701 TTCCAGATGAGTATGGAAATACCACTCTGCACTACGCTATCTATAATGAA 750

     251 AspLysLeuMetAlaLysAlaLeuLeuLeuTyrGlyAlaAspIleGluSe 267
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     751 GATAAATTAATGGCCAAAGCACTGCTCTTATATGGTGCTGATATCGAATC 800

     267 rLysAsnLysHisGlyLeuThrProLeuLeuLeuGlyValHisGluGlnL 284
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     801 AAAAAACAAGCATGGCCTCACACCACTGTTACTTGGTGTACATGAGCAAA 850

     284 ysGlnGlnValValLysPheLeuIleLysLysLysAlaAsnLeuAsnAla 300
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     851 AACAGCAAGTCGTGAAATTTTTAATCAAGAAAAAAGCGAATTTAAATGCA 900

     301 LeuAspArgTyrGlyArgThrAlaLeuIleLeuAlaValCysCysGlySe 317
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     901 CTGGATAGATATGGAAGGACTGCTCTCATACTTGCTGTATGTTGTGGATC 950

     317 rAlaSerIleValSerLeuLeuLeuGluGlnAsnIleAspValSerSerG 334
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     951 AGCAAGTATAGTCAGCCTTCTACTTGAGCAAAATATTGATGTATCTTCTC 1000

     334 lnAspLeuSerGlyGlnThrAlaArgGluTyrAlaValSerSerHisHis 350
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    1001 AAGATCTATCTGGACAGACGGCCAGAGAGTATGCTGTTTCTAGTCATCAT 1050

     351 HisValIleCysGlnLeuLeuSerAspTyrLysGluLysGlnMetLeuLy 367
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    1051 CATGTAATTTGCCAGTTACTTTCTGACTACAAAGAAAAACAGATGCTAAA 1100

     367 sIleSerSerGluAsnSerAsnProGlu 376
         |||||||||||||||||||||||||||
    1101 AATCTCTTCTGAAAACAGCAATCCAGAA 1128

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAI67212

seq_documentation_block:
ID   AAI67212 standard; cDNA; 2000 BP.
XX
AC   AAI67212;
XX
DT   11-FEB-2002  (first entry)
XX
DE   B305D isoform C splice variant 2 encoding cDNA.
XX
KW   Genetic subtraction; DNA microarray analysis; polymerase chain reaction;
KW   cancer; B305D; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1971
FT                   /*tag=  a
FT                   /product= "B305D isoform C splice variant"
XX
PN   WO200175171-A2.
XX
PD   11-OCT-2001.
XX
PF   02-APR-2001; 2001WO-US10631.
XX
PR   03-APR-2000; 2000US-194241P.
PR   20-JUL-2000; 2000US-219862P.
```

```
PR   27-JUL-2000; 2000US-221300P.
PR   18-DEC-2000; 2000US-256592P.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Houghton RL,  Dillon DC,  Molesh DA,  Xu J,  Zehentner B,  Persing DH;
XX
DR   WPI; 2001-626449/72.
DR   P-PSDB; AAG65977.
XX
PT   Identifying tissue (tumour)-specific polynucleotides overexpressed in
PT   tissue of interest as compared to control tissue, for detecting cancer
PT   cells in patient, comprises DNA microarray analysis or quantitative
PT   polymerase chain reaction -
XX
PS   Claim 4; Page 94; 127pp; English.
XX
CC   The invention relates to identifying tissue-specific polynucleotides (P)
CC   that involves performing a genetic subtraction to identify pool of (P)
CC   from tissue of interest (TI), performing DNA microarray analysis to
CC   identify first subset of polynucleotides (SP1) at least 2-fold over
CC   expressed in TI, and performing quantitative polymerase chain reaction
CC   (PCR) analysis on SP1 to identify second subset of (P). The method is
CC   useful for determining the presence or absence of a cancer cell in a
CC   patient, monitoring the progression of cancer in a patient using a
CC   biological sample such as blood, serum, lymph nodes, bone marrow, sputum,
CC   urine or a tumour biopsy sample. The methods are useful for determining
CC   the presence or absence of or monitoring progression of prostate, breast,
CC   colon, ovarian, lung, head and neck, lymphoma, leukemia, melanoma, liver,
CC   gastric, kidney, bladder, pancreatic or endometrial cancer. The present
CC   sequence represents a cDNA encoding a B305D isoform C splice variant.
XX
SQ   Sequence 2000 BP; 698 A; 388 C; 489 G; 425 T; 0 other;


alignment_scores:
            Quality: 2024.00              Length:    376
              Ratio:  5.383                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-810-936-304 x AAI67212   ..

 Align seg 1/1  to: AAI67212  from: 1  to: 2000

       1 MetValValGluValAspSerMetProAlaAlaSerSerValLysLysPr 17
         |||||||||||||||||||||||||||||||||||||||||||||||||||
       1 ATGGTGGTTGAGGTTGATTCCATGCCGGCTGCCTCTTCTGTGAAGAAGCC 50

      17 oPheGlyLeuArgSerLysMetGlyLysTrpCysCysArgCysPheProC 34
         |||||||||||||||||||||||||||||||||||||||||||||||||||
      51 ATTTGGTCTCAGGAGCAAGATGGGCAAGTGGTGCTGCCGTTGCTTCCCCT 100

      34 ysCysArgGluSerGlyLysSerAsnValGlyThrSerGlyAspHisAsp 50
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     101 GCTGCAGGGAGAGCGGCAAGAGCAACGTGGGCACTTCTGGAGACCACGAC 150

      51 AspSerAlaMetLysThrLeuArgSerLysMetGlyLysTrpCysArgHi 67
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     151 GACTCTGCTATGAAGACACTCAGGAGCAAGATGGGCAAGTGGTGCCGCCA 200

      67 sCysPheProCysCysArgGlySerGlyLysSerAsnValGlyAlaSerG 84
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     201 CTGCTTCCCCTGCTGCAGGGGGAGTGGCAAGAGCAACGTGGGCGCTTCTG 250

      84 lyAspHisAspAspSerAlaMetLysThrLeuArgAsnLysMetGlyLys 100
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     251 GAGACCACGACGACTCTGCTATGAAGACACTCAGGAACAAGATGGGCAAG 300

     101 TrpCysCysHisCysPheProCysCysArgGlySerGlyLysSerLysVa 117
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     301 TGGTGCTGCCACTGCTTCCCCTGCTGCAGGGGGAGCGGCAAGAGCAAGGT 350

     117 lGlyAlaTrpGlyAspTyrAspAspSerAlaPheMetGluProArgTyrH 134
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     351 GGGCGCTTGGGGAGACTACGATGACAGTGCCTTCATGGAGCCCAGGTACC 400

     134 isValArgGlyGluAspLeuAspLysLeuHisArgAlaAlaTrpTrpGly 150
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     401 ACGTCCGTGGAGAAGATCTGGACAAGCTCCACAGAGCTGCCTGGTGGGGT 450

     151 LysValProArgLysAspLeuIleValMetLeuArgAspThrAspValAs 167
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     451 AAAGTCCCCAGAAAGGATCTCATCGTCATGCTCAGGGACACTGACGTGAA 500

     167 nLysLysAspLysGlnLysArgThrAlaLeuHisLeuAlaSerAlaAsnG 184
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     501 CAAGAAGGACAAGCAAAAGAGGACTGCTCTACATCTGGCCTCTGCCAATG 550

     184 lyAsnSerGluValValLysLeuLeuLeuAspArgArgCysGlnLeuAsn 200
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     551 GGAATTCAGAAGTAGTAAAACTCCTGCTGGACAGACGATGTCAACTTAAT 600

     201 ValLeuAspAsnLysLysArgThrAlaLeuIleLysAlaValGlnCysGl 217
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     601 GTCCTTGACAACAAAAAGAGGACAGCTCTGATAAAGGCCGTACAATGCCA 650

     217 nGluAspGluCysAlaLeuMetLeuLeuGluHisGlyThrAspProAsnI 234
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     651 GGAAGATGAATGTGCGTTAATGTTGCTGGAACATGGCACTGATCCAAATA 700

     234 leProAspGluTyrGlyAsnThrThrLeuHisTyrAlaIleTyrAsnGlu 250
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     701 TTCCAGATGAGTATGGAAATACCACTCTGCACTACGCTATCTATAATGAA 750

     251 AspLysLeuMetAlaLysAlaLeuLeuLeuTyrGlyAlaAspIleGluSe 267
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     751 GATAAATTAATGGCCAAAGCACTGCTCTTATATGGTGCTGATATCGAATC 800

     267 rLysAsnLysHisGlyLeuThrProLeuLeuLeuGlyValHisGluGlnL 284
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     801 AAAAAACAAGCATGGCCTCACACCACTGTTACTTGGTGTACATGAGCAAA 850

     284 ysGlnGlnValValLysPheLeuIleLysLysLysAlaAsnLeuAsnAla 300
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     851 AACAGCAAGTCGTGAAATTTTTAATCAAGAAAAAAGCGAATTTAAATGCA 900

     301 LeuAspArgTyrGlyArgThrAlaLeuIleLeuAlaValCysCysGlySe 317
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     901 CTGGATAGATATGGAAGGACTGCTCTCATACTTGCTGTATGTTGTGGATC 950

     317 rAlaSerIleValSerLeuLeuLeuGluGlnAsnIleAspValSerSerG 334
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     951 AGCAAGTATAGTCAGCCTTCTACTTGAGCAAAATATTGATGTATCTTCTC 1000

     334 lnAspLeuSerGlyGlnThrAlaArgGluTyrAlaValSerSerHisHis 350
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    1001 AAGATCTATCTGGACAGACGGCCAGAGAGTATGCTGTTTCTAGTCATCAT 1050

     351 HisValIleCysGlnLeuLeuSerAspTyrLysGluLysGlnMetLeuLy 367
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    1051 CATGTAATTTGCCAGTTACTTTCTGACTACAAAGAAAAACAGATGCTAAA 1100

     367 sIleSerSerGluAsnSerAsnProGlu 376
         |||||||||||||||||||||||||||||
    1101 AATCTCTTCTGAAAACAGCAATCCAGAA 1128

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAS63808

seq_documentation_block:
ID   AAS63808 standard; cDNA; 2000 BP.
XX
AC   AAS63808;
XX
DT   29-JAN-2002  (first entry)
XX
```

```
DE   Human prostate cDNA clone B305D splice variant #9.
XX
KW   Human; prostate cancer; ss; cytostatic; immunostimulant; tumour.
XX
OS   Homo sapiens.
XX
PN   WO200173032-A2.
XX
PD   04-OCT-2001.
XX
PF   27-MAR-2001; 2001WO-US09919.
XX
PR   27-MAR-2000; 2000US-0536857.
PR   09-MAY-2000; 2000US-0568100.
PR   12-MAY-2000; 2000US-0570737.
PR   13-JUN-2000; 2000US-0593793.
PR   27-JUN-2000; 2000US-0605783.
PR   10-AUG-2000; 2000US-0636215.
PR   29-AUG-2000; 2000US-0651236.
PR   06-SEP-2000; 2000US-0657279.
PR   02-OCT-2000; 2000US-0679426.
PR   10-OCT-2000; 2000US-0685166.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD;
PI   Fanger GR, Retter MW, Stolk JA, Day CH, Vedvick TS, Carter D;
PI   Li SX, Wang A, Skeiky YAW, Hepler WT, Henderson RA;
XX
DR   WPI; 2001-639232/73.
DR   P-PSDB; AAU69778.
XX
PT   New human prostate-specific polypeptides and polynucleotides useful for
PT   the diagnosis and treatment of cancer, especially prostate cancer -
XX
PS   Claim 1; Page 349-350; 579pp; English.
XX
CC   The invention relates to isolated prostate-specific
CC   polynucleotides, polypeptides, fusion proteins of the polypeptides,
CC   antibodies raised against the polypeptides (or antigenic epitopes
CC   derived from them) and antigen-presenting cells expressing the
CC   polypeptides. The antibodies are useful for detecting the presence of
CC   cancer, especially prostate cancer. The polypeptides, polynucleotides and
CC   the antigen-presenting cells are useful for stimulating and/or expanding
CC   T cells specific for a tumour protein, and for inhibiting the development
CC   of cancer especially prostate cancer. Compositions comprising the
CC   polynucleotide and/or polypeptide are useful for stimulating an immune
CC   response, and for treating cancer. The oligonucleotide is useful for
CC   detecting cancer. The present sequence is a prostate specific
CC   polynucleotide of the invention.
XX
SQ   Sequence 2000 BP; 698 A; 388 C; 489 G; 425 T; 0 other;


alignment_scores:
            Quality: 2024.00              Length:    376
              Ratio:  5.383                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-810-936-304 x AAS63808   ..

 Align seg 1/1  to: AAS63808  from: 1  to: 2000

       1 MetValValGluValAspSerMetProAlaAlaSerSerValLysLysPr 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 ATGGTGGTTGAGGTTGATTCCATGCCGGCTGCCTCTTCTGTGAAGAAGCC 50

      17 oPheGlyLeuArgSerLysMetGlyLysTrpCysCysArgCysPheProC 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 ATTTGGTCTCAGGAGCAAGATGGGCAAGTGGTGCTGCCGTTGCTTCCCCT 100

      34 ysCysArgGluSerGlyLysSerAsnValGlyThrSerGlyAspHisAsp 50
                                                              ||||||||||||||||||||||||||||||||||||||||||||||||||
                                                            101 GCTGCAGGGAGAGCGGCAAGAGCAACGTGGGCACTTCTGGAGACCACGAC 150

                                                             51 AspSerAlaMetLysThrLeuArgSerLysMetGlyLysTrpCysArgHi 67
                                                                ||||||||||||||||||||||||||||||||||||||||||||||||||
                                                            151 GACTCTGCTATGAAGACACTCAGGAGCAAGATGGGCAAGTGGTGCCGCCA 200

                                                             67 sCysPheProCysCysArgGlySerGlyLysSerAsnValGlyAlaSerG 84
                                                                ||||||||||||||||||||||||||||||||||||||||||||||||||
                                                            201 CTGCTTCCCCTGCTGCAGGGGGAGTGGCAAGAGCAACGTGGGCGCTTCTG 250

                                                             84 lyAspHisAspAspSerAlaMetLysThrLeuArgAsnLysMetGlyLys 100
                                                                ||||||||||||||||||||||||||||||||||||||||||||||||||
                                                            251 GAGACCACGACGACTCTGCTATGAAGACACTCAGGAACAAGATGGGCAAG 300

                                                            101 TrpCysCysHisCysPheProCysCysArgGlySerGlyLysSerLysVa 117
                                                                ||||||||||||||||||||||||||||||||||||||||||||||||||
                                                            301 TGGTGCTGCCACTGCTTCCCCTGCTGCAGGGGGAGCGGCAAGAGCAAGGT 350

                                                            117 lGlyAlaTrpGlyAspTyrAspAspSerAlaPheMetGluProArgTyrH 134
                                                                ||||||||||||||||||||||||||||||||||||||||||||||||||
                                                            351 GGGCGCTTGGGGAGACTACGATGACAGTGCCTTCATGGAGCCCAGGTACC 400

                                                            134 isValArgGlyGluAspLeuAspLysLeuHisArgAlaAlaTrpTrpGly 150
                                                                ||||||||||||||||||||||||||||||||||||||||||||||||||
                                                            401 ACGTCCGTGGAGAAGATCTGGACAAGCTCCACAGAGCTGCCTGGTGGGGT 450

                                                            151 LysValProArgLysAspLeuIleValMetLeuArgAspThrAspValAs 167
                                                                ||||||||||||||||||||||||||||||||||||||||||||||||||
                                                            451 AAAGTCCCCAGAAAGGATCTCATCGTCATGCTCAGGGACACTGACGTGAA 500

                                                            167 nLysLysAspLysGlnLysArgThrAlaLeuHisLeuAlaSerAlaAsnG 184
                                                                ||||||||||||||||||||||||||||||||||||||||||||||||||
                                                            501 CAAGAAGGACAAGCAAAAGAGGACTGCTCTACATCTGGCCTCTGCCAATG 550

                                                            184 lyAsnSerGluValValLysLeuLeuLeuAspArgArgCysGlnLeuAsn 200
                                                                ||||||||||||||||||||||||||||||||||||||||||||||||||
                                                            551 GGAATTCAGAAGTAGTAAAACTCCTGCTGGACAGACGATGTCAACTTAAT 600

                                                            201 ValLeuAspAsnLysLysArgThrAlaLeuIleLysAlaValGlnCysGl 217
                                                                ||||||||||||||||||||||||||||||||||||||||||||||||||
                                                            601 GTCCTTGACAACAAAAAGAGGACAGCTCTGATAAAGGCCGTACAATGCCA 650

                                                            217 nGluAspGluCysAlaLeuMetLeuLeuGluHisGlyThrAspProAsnI 234
                                                                ||||||||||||||||||||||||||||||||||||||||||||||||||
                                                            651 GGAAGATGAATGTGCGTTAATGTTGCTGGAACATGGCACTGATCCAAATA 700

                                                            234 leProAspGluTyrGlyAsnThrThrLeuHisTyrAlaIleTyrAsnGlu 250
                                                                ||||||||||||||||||||||||||||||||||||||||||||||||||
                                                            701 TTCCAGATGAGTATGGAAATACCACTCTGCACTACGCTATCTATAATGAA 750

                                                            251 AspLysLeuMetAlaLysAlaLeuLeuLeuTyrGlyAlaAspIleGluSe 267
                                                                ||||||||||||||||||||||||||||||||||||||||||||||||||
                                                            751 GATAAATTAATGGCCAAAGCACTGCTCTTATATGGTGCTGATATCGAATC 800

                                                            267 rLysAsnLysHisGlyLeuThrProLeuLeuLeuGlyValHisGluGlnL 284
                                                                ||||||||||||||||||||||||||||||||||||||||||||||||||
                                                            801 AAAAAACAAGCATGGCCTCACACCACTGTTACTTGGTGTACATGAGCAAA 850

                                                            284 ysGlnGlnValValLysPheLeuIleLysLysLysAlaAsnLeuAsnAla 300
                                                                ||||||||||||||||||||||||||||||||||||||||||||||||||
                                                            851 AACAGCAAGTCGTGAAATTTTTAATCAAGAAAAAAGCGAATTTAAATGCA 900

                                                            301 LeuAspArgTyrGlyArgThrAlaLeuIleLeuAlaValCysCysGlySe 317
                                                                ||||||||||||||||||||||||||||||||||||||||||||||||||
                                                            901 CTGGATAGATATGGAAGGACTGCTCTCATACTTGCTGTATGTTGTGGATC 950

                                                            317 rAlaSerIleValSerLeuLeuLeuGluGlnAsnIleAspValSerSerG 334
                                                                ||||||||||||||||||||||||||||||||||||||||||||||||||
                                                            951 AGCAAGTATAGTCAGCCTTCTACTTGAGCAAAATATTGATGTATCTTCTC 1000

                                                            334 lnAspLeuSerGlyGlnThrAlaArgGluTyrAlaValSerSerHisHis 350
                                                                ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
    1001 AAGATCTATCTGGACAGACGGCCAGAGAGTATGCTGTTTCTAGTCATCAT 1050

     351 HisValIleCysGlnLeuLeuSerAspTyrLysGluLysGlnMetLeuLy 367
         |||||||||||||||||||||||||||||||||||||||||||||||||||
    1051 CATGTAATTTGCCAGTTACTTTCTGACTACAAAGAAAAACAGATGCTAAA 1100

     367 sIleSerSerGluAsnSerAsnProGlu 376
         |||||||||||||||||||||||||||
    1101 AATCTCTTCTGAAAACAGCAATCCAGAA 1128

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAH93715

seq_documentation_block:
ID   AAH93715 standard; cDNA; 2000 BP.
XX
AC   AAH93715;
XX
DT   04-OCT-2001  (first entry)
XX
DE   Human prostate-specific cDNA sequence B305D splice variant #9.
XX
KW   Human; prostate cancer; prostate-specific; diagnosis; vaccine;
KW   cytostatic; gene therapy; metastasis; ss.
XX
OS   Homo sapiens.
XX
PN   WO200151633-A2.
XX
PD   19-JUL-2001.
XX
PF   16-JAN-2001; 2001WO-US01574.
XX
PR   14-JAN-2000; 2000US-0483672.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Xu J,  Dillon DC,  Mitcham JL,  Harlocker SL,  Jiang Y,  Reed SG;
PI   Kalos MD,  Fanger GR,  Day CH,  Retter MW,  Stolk JA,  Skeiky YAW;
PI   Wang A,  Meagher MJ;
XX
DR   WPI; 2001-425873/45.
XX
PT   New polynucleotide encoding a prostate-specific protein, for
PT   diagnosing, monitoring and treating prostate cancer in a patient and
PT   for use in vaccines -
XX
PS   Claim 1; Page 347-348; 543pp; English.
XX
CC   The present invention describes polynucleotide sequences (I) which encode
CC   prostate-specific proteins (II). (I) and (II) have cytostatic activity,
CC   and can be used in vaccine production and gene therapy. (I), (II),
CC   antibodies to (II), fusion proteins comprising (II), and isolated
CC   T cells prepared using (I) or (II) are used treat cancer in a patient.
CC   (I) and the antibodies are also used in the detection of cancer in a
CC   patient. The cancer that is diagnosed or treated is particularly
CC   prostate cancer. (I) and (II) can be used in vaccines. The antibodies or
CC   (I) can be used for monitoring the progression of cancer in a patient.
CC   (I) and (II) can also be used to improve diagnostic and therapeutic
CC   methods for prostate cancer. They can indicate the level of metastasis
CC   as well as the prostate volume. AAH93357 to AAH93944 and AAM01115 to
CC   AAM01318 represent polynucleotide and amino acid sequences used in the
CC   exemplification of the present invention.
XX
SQ   Sequence 2000 BP; 698 A; 388 C; 489 G; 425 T; 0 other;


alignment_scores:
            Quality: 2024.00              Length:    376
              Ratio:  5.383                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-810-936-304 x AAH93715   ..
```

```
Align seg 1/1  to: AAH93715  from: 1  to: 2000

      1 MetValValGluValAspSerMetProAlaAlaSerSerValLysLysPr 17
        ||||||||||||||||||||||||||||||||||||||||||||||||||
      1 ATGGTGGTTGAGGTTGATTCCATGCCGGCTGCCTCTTCTGTGAAGAAGCC 50

     17 oPheGlyLeuArgSerLysMetGlyLysTrpCysCysArgCysPheProC 34
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     51 ATTTGGTCTCAGGAGCAAGATGGGCAAGTGGTGCTGCCGTTGCTTCCCCT 100

     34 ysCysArgGluSerGlyLysSerAsnValGlyThrSerGlyAspHisAsp 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    101 GCTGCAGGGAGAGCGGCAAGAGCAACGTGGGCACTTCTGGAGACCACGAC 150

     51 AspSerAlaMetLysThrLeuArgSerLysMetGlyLysTrpCysArgHi 67
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    151 GACTCTGCTATGAAGACACTCAGGAGCAAGATGGGCAAGTGGTGCCGCCA 200

     67 sCysPheProCysCysArgGlySerGlyLysSerAsnValGlyAlaSerG 84
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    201 CTGCTTCCCCTGCTGCAGGGGGAGTGGCAAGAGCAACGTGGGCGCTTCTG 250

     84 lyAspHisAspAspSerAlaMetLysThrLeuArgAsnLysMetGlyLys 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    251 GAGACCACGACGACTCTGCTATGAAGACACTCAGGAACAAGATGGGCAAG 300

    101 TrpCysCysHisCysPheProCysCysArgGlySerGlyLysSerLysVa 117
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    301 TGGTGCTGCCACTGCTTCCCCTGCTGCAGGGGGAGCGGCAAGAGCAAGGT 350

    117 lGlyAlaTrpGlyAspTyrAspAspSerAlaPheMetGluProArgTyrH 134
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    351 GGGCGCTTGGGGAGACTACGATGACAGTGCCTTCATGGAGCCCAGGTACC 400

    134 isValArgGlyGluAspLeuAspLysLeuHisArgAlaAlaTrpTrpGly 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    401 ACGTCCGTGGAGAAGATCTGGACAAGCTCCACAGAGCTGCCTGGTGGGGT 450

    151 LysValProArgLysAspLeuIleValMetLeuArgAspThrAspValAs 167
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    451 AAAGTCCCCAGAAAGGATCTCATCGTCATGCTCAGGGACACTGACGTGAA 500

    167 nLysLysAspLysGlnLysArgThrAlaLeuHisLeuAlaSerAlaAsnG 184
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    501 CAAGAAGGACAAGCAAAAGAGGACTGCTCTACATCTGGCCTCTGCCAATG 550

    184 lyAsnSerGluValValLysLeuLeuLeuAspArgArgCysGlnLeuAsn 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    551 GGAATTCAGAAGTAGTAAAACTCCTGCTGGACAGACGATGTCAACTTAAT 600

    201 ValLeuAspAsnLysLysArgThrAlaLeuIleLysAlaValGlnCysGl 217
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    601 GTCCTTGACAACAAAAAGAGGACAGCTCTGATAAAGGCCGTACAATGCCA 650

    217 nGluAspGluCysAlaLeuMetLeuLeuGluHisGlyThrAspProAsnI 234
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    651 GGAAGATGAATGTGCGTTAATGTTGCTGGAACATGGCACTGATCCAAATA 700

    234 leProAspGluTyrGlyAsnThrThrLeuHisTyrAlaIleTyrAsnGlu 250
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    701 TTCCAGATGAGTATGGAAATACCACTCTGCACTACGCTATCTATAATGAA 750

    251 AspLysLeuMetAlaLysAlaLeuLeuLeuTyrGlyAlaAspIleGluSe 267
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    751 GATAAATTAATGGCCAAAGCACTGCTCTTATATGGTGCTGATATCGAATC 800

    267 rLysAsnLysHisGlyLeuThrProLeuLeuLeuGlyValHisGluGlnL 284
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    801 AAAAAACAAGCATGGCCTCACACCACTGTTACTTGGTGTACATGAGCAAA 850

    284 ysGlnGlnValValLysPheLeuIleLysLysLysAlaAsnLeuAsnAla 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
     851 AACAGCAAGTCGTGAAATTTTTAATCAAGAAAAAAGCGAATTTAAATGCA 900

     301 LeuAspArgTyrGlyArgThrAlaLeuIleLeuAlaValCysCysGlySe 317
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     901 CTGGATAGATATGGAAGGACTGCTCTCATACTTGCTGTATGTTGTGGATC 950

     317 rAlaSerIleValSerLeuLeuLeuGluGlnAsnIleAspValSerSerG 334
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     951 AGCAAGTATAGTCAGCCTTCTACTTGAGCAAAATATTGATGTATCTTCTC 1000

     334 lnAspLeuSerGlyGlnThrAlaArgGluTyrAlaValSerSerHisHis 350
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1001 AAGATCTATCTGGACAGACGGCCAGAGAGTATGCTGTTTCTAGTCATCAT 1050

     351 HisValIleCysGlnLeuLeuSerAspTyrLysGluLysGlnMetLeuLy 367
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1051 CATGTAATTTGCCAGTTACTTTCTGACTACAAAGAAAAACAGATGCTAAA 1100

     367 sIleSerSerGluAsnSerAsnProGlu 376
         |||||||||||||||||||||||||||
    1101 AATCTCTTCTGAAAACAGCAATCCAGAA 1128
```

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAH85029

seq_documentation_block:

ID AAH85029 standard; cDNA; 2000 BP.
XX
AC AAH85029;
XX
DT 25-SEP-2001 (first entry)
XX
DE Human prostate-specific cDNA sequence B305D splice variant #9.
XX
KW Human; prostate cancer; therapy; diagnosis; cat eye syndrome;
KW chromosome 22q11.2; prostate-specific protein; chromosome 1;
KW prostate specific antigen; PSA; ss.
XX
OS Homo sapiens.
XX
PN WO200134802-A2.
XX
PD 17-MAY-2001.
XX
PF 09-NOV-2000; 2000WO-US30904.
XX
PR 12-NOV-1999; 99US-0439313.
PR 18-NOV-1999; 99US-0443686.
XX
PA (CORI-) CORIXA CORP.
XX
PI Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Reed SG;
PI Kalos MD, Retter MW, Stolk JA, Day CH, Skeiky YAW, Wang A;
XX
DR WPI; 2001-308785/32.
XX
PT Isolated polypeptide comprising at least an immunogenic portion of a
PT prostate-specific protein, useful in the diagnosis and therapy of
PT prostate cancer -
XX
PS Claim 31; Page 247-248; 325pp; English.
XX
CC The present invention describes an isolated polypeptide (P1) comprising
CC at least an immunogenic portion of a prostate-specific protein, or its
CC variant. Also described are polynucleotides (N1) encoding (P1). (P1) and
CC (N1) have cytostatic activity and can be used in vaccine production.
CC The polypeptides, nucleic acids and antibodies from the present
CC invention are useful in the diagnosis and therapy of prostate cancer.
CC Prostate specific genes P704P, P712P, P774P, P775P and B305D are located
CC in a genomic region on chromosome 22q11.2 known as the Cat Eye Syndrome
CC region. Prostate specific antigen (PSA) P501S was located on
CC chromosome 1. AAH84671 to AAH85143 and AAG99000 to AAG99077 represent
CC polynucleotide and polypeptide sequences used in the exemplification
CC of the present invention.

XX
SQ Sequence 2000 BP; 698 A; 388 C; 489 G; 425 T; 0 other;

alignment_scores:
Quality: 2024.00 Length: 376
Ratio: 5.383 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-810-936-304 x AAH85029 ..

Align seg 1/1 to: AAH85029 from: 1 to: 2000

```
       1 MetValValGluValAspSerMetProAlaAlaSerSerValLysLysPr 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 ATGGTGGTTGAGGTTGATTCCATGCCGGCTGCCTCTTCTGTGAAGAAGCC 50

      17 oPheGlyLeuArgSerLysMetGlyLysTrpCysCysArgCysPheProC 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 ATTTGGTCTCAGGAGCAAGATGGGCAAGTGGTGCTGCCGTTGCTTCCCCT 100

      34 ysCysArgGluSerGlyLysSerAsnValGlyThrSerGlyAspHisAsp 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 GCTGCAGGGAGAGCGGCAAGAGCAACGTGGGCACTTCTGGAGACCACGAC 150

      51 AspSerAlaMetLysThrLeuArgSerLysMetGlyLysTrpCysArgHi 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     151 GACTCTGCTATGAAGACACTCAGGAGCAAGATGGGCAAGTGGTGCCGCCA 200

      67 sCysProCysCysArgGlySerGlyLysSerAsnValGlyAlaSerG 84
         ||||||||||||||||||||||||||||||||||||||||||||||||
     201 CTGCTTCCCCTGCTGCAGGGGGAGTGGCAAGAGCAACGTGGGCGCTTCTG 250

      84 lyAspHisAspAspSerAlaMetLysThrLeuArgAsnLysMetGlyLys 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     251 GAGACCACGACGACTCTGCTATGAAGACACTCAGGAACAAGATGGGCAAG 300

     101 TrpCysCysHisCysProCysCysArgGlySerGlyLysSerLysVa 117
         ||||||||||||||||||||||||||||||||||||||||||||||||
     301 TGGTGCTGCCACTGCTTCCCCTGCTGCAGGGGGAGCGGCAAGAGCAAGGT 350

     117 lGlyAlaTrpGlyAspTyrAspAspSerAlaPheMetGluProArgTyrH 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     351 GGGCGCTTGGGGAGACTACGATGACAGTGCCTTCATGGAGCCCAGGTACC 400

     134 isValArgGlyGluAspLeuAspLysLeuHisArgAlaAlaTrpTrpGly 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     401 ACGTCCGTGGAGAAGATCTGGACAAGCTCCACAGAGCTGCCTGGTGGGGT 450

     151 LysValProArgLysAspLeuIleValMetLeuArgAspThrAspValAs 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     451 AAAGTCCCCAGAAAGGATCTCATCGTCATGCTCAGGGACACTGACGTGAA 500

     167 nLysLysAspLysGlnLysArgThrAlaLeuHisLeuAlaSerAlaAsnG 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     501 CAAGAAGGACAAGCAAAAGAGGACTGCTCTACATCTGGCCTCTGCCAATG 550

     184 lyAsnSerGluValValLysLeuLeuLeuAspArgArgCysGlnLeuAsn 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     551 GGAATTCAGAAGTAGTAAAACTCCTGCTGGACAGACGATGTCAACTTAAT 600

     201 ValLeuAspAsnLysLysArgThrAlaLeuIleLysAlaValGlnCysGl 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     601 GTCCTTGACAACAAAAAGAGGACAGCTCTGATAAAGGCCGTACAATGCCA 650

     217 nGluAspGluCysAlaLeuMetLeuLeuGluHisGlyThrAspProAsnI 234
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     651 GGAAGATGAATGTGCGTTAATGTTGCTGGAACATGGCACTGATCCAAATA 700

     234 leProAspGluTyrGlyAsnThrThrLeuHisTyrAlaIleTyrAsnGlu 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     701 TTCCAGATGAGTATGGAAATACCACTCTGCACTACGCTATCTATAATGAA 750
```

```
 251 AspLysLeuMetAlaLysAlaLeuLeuLeuTyrGlyAlaAspIleGluSe 267
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 751 GATAAATTAATGGCCAAAGCACTGCTCTTATATGGTGCTGATATCGAATC 800

 267 rLysAsnLysHisGlyLeuThrProLeuLeuLeuGlyValHisGluGlnL 284
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 801 AAAAAACAAGCATGGCCTCACACCACTGTTACTTGGTGTACATGAGCAAA 850

 284 ysGlnGlnValValLysPheLeuIleLysLysLysAlaAsnLeuAsnAla 300
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 851 AACAGCAAGTCGTGAAATTTTTAATCAAGAAAAAAGCGAATTTAAATGCA 900

 301 LeuAspArgTyrGlyArgThrAlaLeuIleLeuAlaValCysCysGlySe 317
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 901 CTGGATAGATATGGAAGGACTGCTCTCATACTTGCTGTATGTTGTGGATC 950

 317 rAlaSerIleValSerLeuLeuLeuGluGlnAsnIleAspValSerSerG 334
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 951 AGCAAGTATAGTCAGCCTTCTACTTGAGCAAAATATTGATGTATCTTCTC 1000

 334 lnAspLeuSerGlyGlnThrAlaArgGluTyrAlaValSerSerHisHis 350
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1001 AAGATCTATCTGGACAGACGGCCAGAGAGTATGCTGTTTCTAGTCATCAT 1050

 351 HisValIleCysGlnLeuLeuSerAspTyrLysGluLysGlnMetLeuLy 367
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1051 CATGTAATTTGCCAGTTACTTTCTGACTACAAAGAAAAACAGATGCTAAA 1100

 367 sIleSerSerGluAsnSerAsnProGlu 376
     ||||||||||||||||||||||||||||
1101 AATCTCTTCTGAAAACAGCAATCCAGAA 1128
```

```
OM of: US-09-810-936-304 to: EST:*   out_format : pfs

Date: Aug 26, 2002  3:55 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlh
 -Q=/cgn2_1/USPTO_spool/US09810936/runat_23082002_145849_26159/app_query.fasta_1.450
 -DB=EST -QFMT=fastap -SUFFIX=p2n.rst -GAPOP=12.000 -GAPEXT=4.000
 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
 -DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
 -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZE=500
 -MINLEN=0 -MAXLEN=2000000000 -USER=US09810936_@CGN1_1_2747
 -NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30
 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-810-936-304
Query length: 384
Database: EST:*
Database sequences: 13736207
Database length: -1841457050
Search time (sec): 1603.050000



score_list:
Sequence          Strd Orig    ZScore    EScore Len  ! Documentation  ..
gb_est2:BF676987    +  893.00 1579.99  9.3e-79   865  ! BF676987 602084215F1 NIH_MGC_83
gb_est2:BM469654    +  597.00 1048.69  3.6e-49  1020  ! BM469654 AGENCOURT_6444673 NIH
gb_est1:AU120666    +  587.50 1033.71  2.5e-48   843  ! AU120666 AU120666 HEMBB1 Homo s
gb_est2:BI871077    +  538.00  946.34  1.8e-43   754  ! BI871077 603394250F1 NIH_MGC_90
gb_htc:AK017783     +  521.50  907.85  2.5e-41  1758  ! AK017783 Mus musculus 8 days e
gb_htc:AK015948     +  517.50  906.11  3.2e-41  1054  ! AK015948 Mus musculus adult ma
gb_est2:BG720647    +  496.00  872.08  2.5e-39   694  ! BG720647 602692528F1 NIH_MGC_97
gb_htc:AK015165     +  493.50  860.17  1.2e-38  1400  ! AK015165 Mus musculus adult ma
gb_htc:AK015473     +  493.50  859.86  1.2e-38  1441  ! AK015473 Mus musculus adult ma
gb_htc:AK015012     +  493.50  859.75  1.2e-38  1456  ! AK015012 Mus musculus adult ma
gb_est2:BI830348    +  475.00  832.28  4.1e-37   856  ! BI830348 603073084F1 NIH_MGC_11
gb_est2:BI827177    +  471.00  825.48  9.9e-37   828  ! BI827177 603077385F1 NIH_MGC_11
gb_htc:AK005925     +  471.00  821.47  1.6e-36  1209  ! AK005925 Mus musculus adult ma
gb_est1:AA910780    +  468.00  824.66  1.1e-36   539  ! AA910780 ol48b04.s1 Soares_NFL_
gb_est2:BG716974    +  468.00  819.04  2.2e-36   916  ! BG716974 602689005F1 NIH_MGC_97
gb_est2:BI546858    +  465.00  813.87  4.4e-36   899  ! BI546858 603189764F1 NIH_MGC_95
gb_est2:BI828139    +  465.00  813.20  4.8e-36   958  ! BI828139 603073604F1 NIH_MGC_11
gb_est2:BI831391    +  464.00  812.50  5.2e-36   864  ! BI831391 603074383F1 NIH_MGC_11
gb_est2:BG703856    +  463.00  811.40  6.0e-36   810  ! BG703856 602686889F1 NIH_MGC_95
gb_est2:BG721483    +  462.00  808.41  8.8e-36   907  ! BG721483 602695061F1 NIH_MGC_97
gb_est2:BI826430    +  460.00  806.95  1.1e-35   743  ! BI826430 603076109F1 NIH_MGC_11
gb_est2:BI560539    +  456.00  799.26  2.8e-35   781  ! BI560539 603254543F1 NIH_MGC_97
gb_est2:BI549417    +  456.00  798.90  3.0e-35   808  ! BI549417 603190072F1 NIH_MGC_95
gb_est2:BI831462    +  456.00  798.09  3.3e-35   872  ! BI831462 603074566F1 NIH_MGC_11
gb_est2:BI831866    +  454.00  796.14  4.2e-35   748  ! BI831866 603078713F1 NIH_MGC_11
gb_est2:BI520473    +  454.00  795.04  4.9e-35   830  ! BI520473 603071669F1 NIH_MGC_11
gb_est2:BI825921    +  453.50  795.48  4.6e-35   732  ! BI825921 603076876F1 NIH_MGC_11
gb_htc:AK016868     +  447.00  778.58  4.0e-34  1203  ! AK016868 Mus musculus adult ma
gb_htc:AK016475     +  447.00  774.13  7.1e-34  1832  ! AK016475 Mus musculus adult ma
gb_est2:BI462821    +  445.50  779.26  3.7e-34   876  ! BI462821 603205786F1 NIH_MGC_97
gb_htc:AK015247     +  445.00  770.52  1.1e-33  1837  ! AK015247 Mus musculus adult ma
gb_est2:BG719053    +  437.00  765.14  2.3e-33   791  ! BG719053 602699174F1 NIH_MGC_97
gb_est2:BF675057    +  435.00  761.15  3.8e-33   822  ! BF675057 602136651F1 NIH_MGC_83
gb_est2:BG702026    +  434.50  760.69  4.0e-33   789  ! BG702026 602683812F1 NIH_MGC_95
gb_est2:BG717440    +  434.00  760.38  4.2e-33   747  ! BG717440 602689791F1 NIH_MGC_97
gb_est2:BG718834    +  427.00  747.75  2.1e-32   754  ! BG718834 602699306F1 NIH_MGC_97
gb_est2:BG717899    +  424.50  742.53  4.1e-32   809  ! BG717899 602693943F1 NIH_MGC_97
gb_est2:BG720794    +  420.00  735.76  9.8e-32   717  ! BG720794 602692016F1 NIH_MGC_97
gb_est2:BI460053    +  416.00  728.26  2.6e-31   741  ! BI460053 603201605F1 NIH_MGC_97
gb_est2:BI827444    +  415.50  727.61  2.8e-31   724  ! BI827444 603077565F1 NIH_MGC_11
gb_est2:BG722705    +  415.00  727.18  2.9e-31   693  ! BG722705 602694719F1 NIH_MGC
gb_est2:BI459125    +  414.50  724.87  4.0e-31   792  ! BI459125 603199496F1 NIH_MGC
gb_est2:BI825466    +  414.00  722.20  5.6e-31   936  ! BI825466 603072210F1 NIH_MGC
gb_est2:BG700835    +  408.00  713.58  1.7e-30   767  ! BG700835 602681736F1 NIH_MGC
gb_est2:BI829067    +  403.50  705.98  4.5e-30   735  ! BI829067 603079158F1 NIH_MGC

seq_name: gb_est2:BF676987

seq_documentation_block:
LOCUS       BF676987               865 bp    mRNA    linear   EST 21-DEC-2000
DEFINITION  602084215F1 NIH_MGC_83 Homo sapiens cDNA clone IMAGE:4248746 5',
            mRNA sequence.
ACCESSION   BF676987
VERSION     BF676987.1  GI:11950882
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 865)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONETECH Laboratories, Inc.
            cDNA Library Preparation: CLONETECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1067  row: m  column: 03
            High quality sequence stop: 642.
FEATURES             Location/Qualifiers
     source          1. .865
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4248746"
                     /clone_lib="NIH_MGC_83"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: prostate; Vector: pDNR-LIB (Clontech);
                     Site_1: SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc
                     ); 5' and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence:  5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                     (where B = A, C, or G and N = A, C,  G, or T).   Average
                     insert size 1.4 kb (range 0.5-4.0 kb).  14/15 colonies
                     contained inserts  by PCR. This library was enriched for
                     full-length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA)."
BASE COUNT      297 a    140 c    211 g    217 t
ORIGIN


alignment_scores:
            Quality:  893.00             Length:    194
              Ratio:  4.880                Gaps:      1
 Percent Similarity:  94.330   Percent Identity: 94.330

alignment_block:
 US-09-810-936-304 x BF676987   ..

 Align seg 1/1  to: BF676987  from: 1  to: 865

     200 AsnValLeuAspAsnLysLysArgThrAlaLeuIleLys..... 212
         |||||||||||||||||||||||||||||||||   |||
       2 AATGTCCTTGACAACAAAAAGAGGACAGCTCTGACAAAGGTATGCAGTAG 51

     213 ..................AlaValGlnCysGlnGluAspGluCysAlaL 223
                            ||||||||||||||||||||||||||||||||
      52 CCAACTATATCAGCATGAGGCCGTACAATGCCAGGAAGATGAATGTGCGT 101
```

```
     223 euMetLeuLeuGluHisGlyThrAspProAsnIleProAspGluTyrGly 239
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     102 TAATGTTGCTGGAACATGGCACTGATCCAAACATTCCAGATGAGTATGGA 151

     240 AsnThrThrLeuHisTyrAlaIleTyrAsnGluAspLysLeuMetAlaLy 256
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     152 AATACCACTCTACACTACGCTATCTACAATGAAGATAAATTAATGGCCAA 201

     256 sAlaLeuLeuLeuTyrGlyAlaAspIleGluSerLysAsnLysHisGlyL 273
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     202 AGCACTGCTCTTATACGGTGCTGATATCGAATCAAAAAACAAGCATGGCC 251

     273 euThrProLeuLeuLeuGlyValHisGluGlnLysGlnGlnValValLys 289
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     252 TCACACCACTGCTACTTGGTGTACATGAGCAAAAACAGCAAGTGGTGAAA 301

     290 PheLeuIleLysLysLysAlaAsnLeuAsnAlaLeuAspArgTyrGlyAr 306
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     302 TTTTTAATCAAGAAAAAAGCGAATTTAAATGCGCTGGATAGATATGGAAG 351

     306 gThrAlaLeuIleLeuAlaValCysCysGlySerAlaSerIleValSerL 323
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     352 AACTGCTCTCATACTTGCTGTATGTTGTGGATCAGCAAGTATAGTCAGCC 401

     323 euLeuLeuGluGlnAsnIleAspValSerSerGlnAspLeuSerGlyGln 339
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     402 TTCTACTTGAGCAAAATATTGATGTATCTTCTCAAGATCTATCTGGACAG 451

     340 ThrAlaArgGluTyrAlaValSerSerHisHisHisValIleCysGlnLe 356
         ||||||||||||||||||||||||||||||||||||    ||||||||
     452 ACGGCCAGAGAGTATGCTGTTTCTAGTCATCATCATGTAAT.TGCCAGTT 500

     356 uLeuSerAspTyrLysGluLysGlnMetLeuLysIleSerSerGluAsnS 373
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
     501 ACTTTCTGACTACAAAGAAAAACAGATGCTAAAAATCTCTTCTGAAAACA 550

     373 erAsnProGluAsnValSerArgThrArgAsn 383
         |||||||||||||||||||||||||||||||
     551 GCAATCCAGAAAATGTCTCAAGAACCAGAAAT 582

seq_name: gb_est2:BM469654

seq_documentation_block:
LOCUS       BM469654               1020 bp    mRNA    linear   EST 05-FEB-2002
DEFINITION  AGENCOURT_6444673 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:5585453
            5', mRNA sequence.
ACCESSION   BM469654
VERSION     BM469654.1  GI:18518696
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1020)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12351  row: m  column: 06
            High quality sequence stop: 680.
FEATURES             Location/Qualifiers
     source          1. .1020
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5585453"
                     /clone_lib="NIH_MGC_92"
                     /tissue_type="embryonal carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: testis; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 2.5 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      348 a    208 c    218 g    240 t      6 others
ORIGIN


alignment_scores:
           Quality:  597.00            Length:    257
             Ratio:  3.411               Gaps:      4
 Percent Similarity:  68.093   Percent Identity: 49.805

alignment_block:
 US-09-810-936-304 x BM469654   ..

 Align seg 1/1  to: BM469654  from: 1  to: 1020

     134 HisValArgGlyGluAspLeuAspLysLeuHisArgAlaAlaTrpTrpGl 150
         |||||||||   :::|||||||   |||:::|||:::||||||      ||
       2 CACGTCCGAGACCGAGATCTCGGCAAGATCCACAAAGCTGCCAGCGCGGG 51

     150 yLysValProArgLysAspLeuIleValMetLeuArgAspThrAspValA 167
         |:::|||   :::   :::   |||:::::|||||   :::   :::|
      52 TAATGTGGCGAAAGTGCAGCAGATCCTTTTGCTCAGGAAGAATGGCTTGA 101

     167 snLysLysAspLysGlnLysArgThrAlaLeuHisLeuAlaSerAlaAsn 183
         ||   :::|||||:::::|||||||||||||||||||||   ||||||
     102 ACGATAGAGACAAGATGAACAGGACGGCTCTACATTTGGCCTGTGCCAAT 151

     184 GlyAsnSerGluValValLysLeuLeuLeuAspArgArgCysGlnLeuAs 200
         |||:::   |||||||||   ||||||:::||||||:::|||||||||||
     152 GGTCATCCAGAAGTAGTAACTCTCCTGGTGGACAGAAAATGCCAGCTCAA 201

     200 nValLeuAspAsnLysLysArgThrAlaLeuIleLysAlaValGlnCysG 217
         ||||   ||||||:::::|||||||||||:::|||||||||||||||||
     202 TGTCTGTGACAACGAAAACAGGACAGCTCTGATGAAGGCTGTACAATGCC 251

     217 lnGluAspGluCysAlaLeuMetLeuLeuGluHisGlyThrAspProAsn 233
         |||||:::::|||||   :::|||||||||||||||:::|||||||||
     252 AGGAAGAGAAATGTGCAACTATTCTGCTAGAACATGGTGCTGATCCAAAT 301

     234 IleProAspGluTyrGlyAsnThrThrLeuHisTyrAlaIleTyrAsnGl 250
         :::   |||   :::|||||||||:::|||||||||||:::|||||||||
     302 CTTGCGGATGTCCATGGCAACACTGCTCTTCACTATGCTGTCTATAATGA 351

     250 uAspLysLeuMetAlaLysAlaLeuLeuLeuTyrGlyAlaAspIleGluS 267
         ||||     :::|||     ||||||||||||   |||:::||||||:
     352 GGACATATCAGTAGCAACAAAGCTGCTTTTGTATGATGCAAATATTGAAG 401

     267 erLysAsnLysHisGlyLeuThrProLeuLeuLeuGlyValHisGluGln 283
         ::|||||||||       ||||||||||||||||||:::|||   |||:::
     402 CAAAAAACAAGGATGACCTCACACCACTTTTACTTGCAGTAAGTGAAAAA 451

     284 LysGlnGlnValValLysPheLeuIleLysLysLysAlaAsnLeuAsnAl 300
         |||||||||:::|||:::||||||||||||||||||||||||:::|||||
     452 AAGCAGCAAATGGTGGAATTTTTAATAAAGAAAAAAGCAAATGTAAATGC 501

     300 aLeuAspArgTyrGlyArgThrAlaLeuIleLeuAlaValCysCysGlyS 317
         |:::|||:::
     502 AGTAGATAAG..................................... 511

     317 erAlaSerIleValSerLeuLeuLeuGluGlnAsnIleAspValSerSer 333
                                  ||||||
     512 .....................TTGGAA..................... 517

     334 GlnAspLeuSerGlyGlnThrAlaArgGluTyrAlaValSerSerHisHi 350
                                                 |||||||||
     518 ....................................AGCAGTCAC.. 526
```

```
     350 sHisValIleCysGlnLeuLeuSerAspTyrLysGluLysGlnMetLeuL 367
                   ||||||:::|||:::|||||||||:::::::::   |
     527 .............CAACTAATTTCAGAATATAAAGAAGAAAGGATACCTA 563

     367 ysIleSerSerGluAsnSerAsn.................ProGluAsn 377
         ||   ||||||:::|||||||||                 ::::::
     564 AACATTCTTCTCAAAATAGTAATTCAGTGGATGAAAGCTCTGAAGACTCC 613

     378 ValSerArgThrArgAsnLys 384
         :::||||||      :::|||
     614 TTAAGCAGGCTTTCTGGCAAA 634

seq_name: gb_est1:AU120666

seq_documentation_block:
LOCUS       AU120666                 843 bp    mRNA    linear   EST 19-OCT-2000
DEFINITION  AU120666 HEMBB1 Homo sapiens cDNA clone HEMBB1001175 5', mRNA
            sequence.
ACCESSION   AU120666
VERSION     AU120666.1  GI:10935901
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 843)
  AUTHORS   Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
            Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
            Isogai,T.
  TITLE     HRI human cDNA project
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Takao Isogai
            Genomics Laboratory
            Helix Research Institute
            1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            Tel: 81-438-52-3951
            Fax: 81-438-52-3952
            Email: genomics@hri.co.jp
            HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
            Research Institute; cDNA library construction: Department of
            Virology, Institute of Medical Science, University of Tokyo, and
            Helix Research Institute.
FEATURES             Location/Qualifiers
     source          1. .843
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="HEMBB1001175"
                     /clone_lib="HEMBB1"
                     /tissue_type="whole embryo, mainly body"
                     /dev_stage="embryo, 10 weeks"
                     /note="Vector: pME18SFL3"
BASE COUNT      240 a    178 c    245 g    176 t      4 others
ORIGIN


alignment_scores:
            Quality:  587.50             Length:    211
              Ratio:  3.539                Gaps:      1
 Percent Similarity:  78.673   Percent Identity: 56.398

alignment_block:
 US-09-810-936-304 x AU120666   ..

 Align seg 1/1  to: AU120666  from: 1  to: 843

      90 AlaMetLysThrLeuArgAsnLysMetGlyLysTrpCysCysHisCysPh 106
         |||||||||   :::  :::|||  |||:::                ||
     164 GCCATGAAGAAGATTTTTAGTAAGAAGGGCGAGTCGCCCTTGGGCTCCTT 213

     106 eProCysCysArgGlySerGlyLysSerLysValGlyAlaTrpGlyAspT 123
         |    |||        :::|||:::::::|||:::   |||:::
     214 CGCGCGGCGGCGG.........AGGAGCAGCGCGGGAGGCGGGGGCGAGC 254

     123 yrAspAspSerAlaPheMetGluProArgTyrHisValArgGlyGluAsp 139
           :::::::|||:::   :::|||   ||||||||||||   :::|||
     255 CGGGGGAGGGCGCCTACTCGCAGCCCGGCTACCACGTCCGAGACCGAGAT 304

     140 LeuAspLysLeuHisArgAlaAlaTrpTrpGlyLysValProArgLysAs 156
         |||   |||:::|||:::||||||      |||:::|||   :::   ::
     305 CTCGGCAAGATCCACAAAGCTGCCAGCGCGGGTAATGTGGCGAAAGTGCA 354

     156 pLeuIleValMetLeuArgAspThrAspValAsnLysLysAspLysGlnL 173
         :   |||:::::::||||||   :::   :::|||   :::|||||::::
     355 GCAGATCCTTTTGCTCAGGAAGAATGGCTTGAACGATAGAGACAAGATGA 404

     173 ysArgThrAlaLeuHisLeuAlaSerAlaAsnGlyAsnSerGluValVal 189
         ::|||||||||||||||||||||   |||||||||:::   |||||||||
     405 ACAGGACGGCTCTACATTTGGCCTGTGCCAATGGTCATCCAGAAGTAGTA 454

     190 LysLeuLeuLeuAspArgArgCysGlnLeuAsnValLeuAspAsnLysLy 206
         ||||||:::||||||:::|||||||||||||||   ||||||:::::
     455 ACTCTCCTGGTGGACAGAAAATGCCAGCTCAATGTCTGTGACAACGAAAA 504

     206 sArgThrAlaLeuIleLysAlaValGlnCysGlnGluAspGluCysAlaL 223
         :|||||||||||:::|||||||||||||||||||||:::::::||||||
     505 CAGGACAGCTCTGATGAAGGCTGTACAATGCCAGGAAGAGAAATGTGCAA 554

     223 euMetLeuLeuGluHisGlyThrAspProAsnIleProAspGluTyrGly 239
           :::|||||||||||||||:::||||||||||:::   |||   :::|||
     555 CTATTCTGCTAGAACATGGTGCTGATCCAAATCTTGCGGATGTCCATGGC 604

     240 AsnThrThrLeuHisTyrAlaIleTyrAsnGluAspLysLeuMetAlaLy 256
         |||||||:::|||||||||||||:::|||||||||||||      :::|||
     605 AACACTGCTCTTCACTATGCTGTCTATAATGAGGACATATCAGTAGCAAC 654

     256 sAlaLeuLeuLeuTyrGlyAlaAspIleGluSerLysAsnLysHisGlyL 273
           ||||||||||||   |||:::||||||:::|||||||||      |
     655 AAAGCTGCTTTTGTATGATGCAAATATTGAAGCAAAAAACAAGGATGACC 704

     273 euThrProLeuLeuLeuGlyValHisGluGlnLysGlnGlnValValLys 289
         |||||||||||||||||:::|||      :::|||||||||:::|||:::
     705 TNACACCACTTTTACTTGCAGTAAGTGGAAAAAAGCAGCAAATGGTGGAA 754

     290 PheLeuIleLysLysLysAlaAsnLeuAsnAla 300
         |||||||||||||||||||||||||:::||||||
     755 TTTTTAATAAAGAAAAAAGCAAATGTAAATGCC 787

seq_name: gb_est2:BI871077

seq_documentation_block:
LOCUS       BI871077                 754 bp    mRNA    linear   EST 11-OCT-2001
DEFINITION  603394250F1 NIH_MGC_90 Homo sapiens cDNA clone IMAGE:5403996 5',
            mRNA sequence.
ACCESSION   BI871077
VERSION     BI871077.1  GI:16044750
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 754)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12030  row: d  column: 13
            High quality sequence stop: 729.
```

```
FEATURES             Location/Qualifiers
     source          1. .754
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5403996"
                     /clone_lib="NIH_MGC_90"
                     /tissue_type="adenocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.7 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      234 a    153 c    171 g    196 t
ORIGIN


alignment_scores:
            Quality:  538.00            Length:    248
              Ratio:  2.924               Gaps:      5
 Percent Similarity:  74.194   Percent Identity: 46.774

alignment_block:
 US-09-810-936-304 x BI871077   ..

 Align seg 1/1  to: BI871077  from: 1  to: 754

     133 TyrHisValArgGlyGluAspLeuAspLysLeuHisArgAlaAlaTrpTr 149
         ||||||::::::|||              :::||||||||::::::::
      10 TATCATCTGAAGGGG..............ATCCACAGAGCTGTCTTCTA 44

     149 pGlyLysValProArgLysAspLeu............IleValMetLeuA 162
         :          :::||||||            :::|||:::|||
      45 T..............CGTGATCTGGAGGAACTGAAGTTCGTTCTGCTCA 79

     162 rgAspThrAspValAsnLysLysAspLysGlnLysArgThrAlaLeuHis 178
               |||:::||||||::|||::::::::||||||||||||||||
      80 CGCGTTATGACATCAATAAGAGAGACAGGAAGGAAAGGACCGCCCTACAT 129

     179 LeuAlaSerAlaAsnGlyAsnSerGluValValLysLeuLeuLeuAspAr 195
         ||||||   |||:::|||:::   |||:::|||   ||||||::::::||
     130 TTGGCCTGTGCCACTGGCCAACCGGAAATGGTACATCTCCTGGTGTCCAG 179

     195 gArgCysGlnLeuAsnValLeuAspAsnLysLysArgThrAlaLeuIleL 212
         |||||||:::|||||||:::   |||::::::   ||||||   |||||||
     180 AAGATGTGAGCTTAACCTCTGCGACCGTGAAGACAGGACACCTCTGATCA 229

     212 ysAlaValGlnCysGlnGluAspGluCysAlaLeuMetLeuLeuGluHis 228
         ||||||||||   :::::::::   ||||||   :::||||||::::::
     230 AGGCTGTACAACTGAGGCAGGAGGCTTGTGCAACTCTTCTGCTGCAAAAT 279

     229 GlyThrAspProAsnIleProAspGluTyrGlyAsnThrThrLeuHisTy 245
         |||:::||||||||||||   |||   :::|||:::|||:::||||||||
     280 GGCGCCGATCCAAATATTACGGATGTCTTTGGAAGGACTGCTCTGCACTA 329

     245 rAlaIleTyrAsnGluAspLysLeuMetAlaLysAlaLeuLeuLeuTyrG 262
         ||||:::||||||||||||      |||   :::   ||||||   ||||
     330 CGCTGTGTATAATGAAGATACATCCATGATAGAAAAACTTCTTTCATATG 379

     262 lyAlaAspIleGluSerLysAsnLysHisGlyLeuThrProLeuLeuLeu 278
         ||||||:::||||||:::   ::::::            ||||||:::|||
     380 GTGCAAATATTGAAGAATGCAGCGAGGATGAATATCCGCCACTGTTCCTT 429

     279 GlyValHisGluGlnLysGlnGlnValValLysPheLeuIleLysLysLy 295
         :::|||   :::::::|||   ::::::|||:::||||||:::||||||||
     430 GCTGTGAGTCAAAGAAAAGTGAAAATGGTGGAATTTTTATTAAAGAAAAA 479

     295 sAlaAsnLeuAsnAlaLeuAspArgTyrGlyArgThrAlaLeuIleLeuA 312
         |||||||:::||||||:::|||      ||||||:::|||||||||   |
     480 AGCAAATATAAATGCCGTTGATTATCTTGGCAGATCAGCCCTCATACATG 529

     312 laValCysCysGlySerAlaSerIleValSerLeuLeuLeuGluGlnAsn 328
         |||||      |||:::   :::||||||   |||||||||:::::|||

     530 CTGTTACTCTTGGAGAAAAAGATATAGTCATTCTTCTTCTGCAGCACAAT 579

     329 IleAspValSerSerGlnAspLeuSerGlyGlnThrAlaArg...GluTy 344
         |||||||||   |||:::               :::   :::   :::||
     580 ATTGATGTGTTTTCTCGAGCATGTGTATGGACAAGCTTGCAGAACGATTA 629

     344 rAlaValSerSerHisHisHisValIleCysGlnLeuLeuSerAspTyrL 361
         ||||   ::::::   ::::::||||||   :::|||:::   ||||||
     630 TGCCAGCGAGGCTAAGAACAGAGTCATTTTTGAACTAATCTATGACTATG 679

     361 ysGluLysGlnMetLeuLysIleSerSerGluAsnSerAsnPro 375
           :::|||:::      ::::::|||      ||||||||||
     680 ACAGAAAGAAACATGAAGAGCTTTCTATA...AATAGCAATCCA 720

seq_name: gb_htc:AK017783

seq_documentation_block:
LOCUS       AK017783               1758 bp    mRNA    linear   HTC 19-JAN-2002
DEFINITION  Mus musculus 8 days embryo whole body cDNA, RIKEN full-length
            enriched library, clone:5730521P14:homolog to to KIAA1074 PROTEIN,
            full insert sequence.
ACCESSION   AK017783
VERSION     AK017783.1  GI:12857204
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) 8 days embryo cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:5730521P14.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2  (sites)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3  (sites)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4  (sites)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 1758)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C.,
            Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T.,
            Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,F.,
            Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T.,
            Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S.,
            Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K.,
            Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J.,
            Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,
            Schriml,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
```

```
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamanaka,I.,
            Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and
            Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Gemomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .1758
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="MGD:MGI:1903054"
                     /db_xref="taxon:10090"
                     /clone="5730521P14"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="8 days embryo"
     CDS             159. .>1758
                     /note="data source:SPTR, source key:Q9UPS8, evidence:ISS
                     homolog to to KIAA1074 PROTEIN
                     putative"
                     /codon_start=1
                     /protein_id="BAB30930.1"
                     /db_xref="GI:12857205"
                     /translation="MKKIFGFRSKGPSPLGPSARPRSNCVGFGRESASGSHVPRYHIH
                     DKDMGKIHKAASVGDVAKVQHILILGKSGVNDRDKKDRTALHLACAYGHPEVVTLLVE
                     RKCEIDARDSESSTALIKAVQCQEEECAAILLDHGADPNVMDSSGNTALHYAVYSENT
                     SMAAKLLAHNANIEAKNKDDLTPMLLAVKENKQHIVEFLVKKKASIHAVDQLGSNRQM
                     FEYDGKRLQRSENSNPVDNGSEDGSLTRSYNTPGPADSWPTSDEEDYNFDNKNVPKIN
                     LTELWTAAQQSRKNQTKCGFEELDNGARFDDSDSPSESEDAIEVEPAPSVRVQTLSPS
                     RQSPDPVEGATELAIEGEENGTDVIESASQEQPNHDNLTRADGWHKSNKSEMMSALGL
                     GEDEDEHSPWDSESISESVSLKDVGHFSGTADQTGKRRAHGQIEDVTYIPSCMSGSRS
                     FKMAKLEESRNVGLPVAHMGMEKNEAPRKYAIMEPTIERRAPVLNKTETVGMTDAQTF
                     KSEPESVSREEQTRLSGSEDSQQRV"
BASE COUNT      531 a    392 c    474 g    361 t
ORIGIN


alignment_scores:
            Quality:  521.50              Length:    386
              Ratio:  2.595                 Gaps:     10
 Percent Similarity:  52.073   Percent Identity: 34.456

alignment_block:
 US-09-810-936-304 x AK017783   ..

 Align seg 1/1  to: AK017783  from: 1  to: 1758

       3 ValGluValAspSerMetProAlaAlaSerSerValLysLysProPheGl 19
         |||   |||   ::::::|||      |||::::::||||||   |||||
     132 GTGCCTGTGAGGGGCGTACCC......TCTGCCATGAAGAAGATTTTTGG 175

      19 yLeuArgSerLysMetGlyLysTrpCysCysArgCysPheProCysCysA 36
         ):::|||||||||
     176 CTTCAGGAGTAAG..................................... 188

      36 rgGluSerGlyLysSerAsnValGlyThrSerGlyAspHisAspAspSer 52
                 |||   |||   :::|||   |||:::
     189 ........GGCCCATCGCCCTTAGGCCCCTCTGCC.............. 215

      53 AlaMetLysThrLeuArgSerLysMetGlyLysTrpCysArgHisCysPh 69
                       |||   :::   :::
     216 ..............CGCCCGCGCAGCAAC..................... 230

      69 eProCysCysArgGlySerGlyLysSerAsnValGlyAlaSerGlyAspH 86
               |||   |||   |||::::::::::      |||||||||:::|
     231 .......TGCGTGGGCTTTGGACGCGAGAGT......GCTAGTGGCTCCC 267

      86 isAspAspSerAlaMetLysThrLeuArgAsnLysMetGlyLysTrpCys 102
         ||
     268 AC..................................................... 269

     103 CysHisCysPheProCysCysArgGlySerGlyLysSerLysValGlyAl 119

     269 ..................................................... 269

     119 aTrpGlyAspTyrAspAspSerAlaPheMetGluProArgTyrHisValA 136
                                           ||||||||||||::::
     270 ............................GTGCCCAGGTACCATATTC 288

     136 rgGlyGluAspLeuAspLysLeuHisArgAlaAlaTrpTrpGlyLysVal 152
         ::   :::|||:::   |||:::|||:::||||||      |||   |||
     289 ACGACAAGGATATGGGCAAGATCCACAAGGCTGCAAGCGTGGGAGATGTA 338

     153 ProArgLysAspLeuIleValMetLeuArgAspThrAspValAsnLysLy 169
            :::   :::   |||:::::::|||     :::   ||||||   ::
     339 GCCAAAGTGCAGCATATCCTTATCCTTGGAAAAAGTGGCGTGAACGATAG 388

     169 sAspLysGlnLysArgThrAlaLeuHisLeuAlaSerAlaAsnGlyAsnS 186
         :|||||:::   ||||||||||||||||||||   |||   |||:::
     389 AGACAAGAAGGACAGGACTGCTCTACATCTTGCCTGTGCTTACGGCCACC 438

     186 erGluValValLysLeuLeuLeuAspArgArgCysGlnLeuAsnValLeu 202
           |||||||||   ||||||:::::|||:::|||:::::::::::::
     439 CAGAAGTGGTGACTCTCTTAGTAGAGAGAAAATGTGAGATAGATGCCCGT 488

     203 AspAsnLysLysArgThrAlaLeuIleLysAlaValGlnCysGlnGluAs 219
         |||:::::::::   ||||||||||||||||||||||||||||||||::
     489 GACAGCGAGAGCAGCACGGCCCTCATTAAGGCCGTGCAGTGCCAGGAGGA 538

     219 pGluCysAlaLeuMetLeuLeuGluHisGlyThrAspProAsnIleProA 236
         :|||||||||   :::||||||:::||||||:::|||||||||:::   |
     539 AGAATGTGCAGCCATACTGCTGGACCACGGGGCCGATCCAAATGTGATGG 588

     236 spGluTyrGlyAsnThrThrLeuHisTyrAlaIleTyrAsnGluAspLys 252
         ||:::   |||||||||:::|||||||||||||:::|||:::|||:::
     589 ACAGCAGCGGCAACACCGCCCTCCACTACGCGGTCTACAGTGAGAACACG 638

     253 LeuMetAlaLysAlaLeuLeuLeuTyrGlyAlaAspIleGluSerLysAs 269
            ||||||      ||||||   ::::::|||:::||||||:::|||||
     639 TCCATGGCAGCCAAACTGCTCGCACACAATGCAAACATAGAAGCCAAAAA 688

     269 nLysHisGlyLeuThrProLeuLeuLeuGlyValHisGluGlnLysGlnG 286
         ||||      |||||||||:::||||||:::|||   |||:::|||||:
     689 CAAGGATGACCTAACACCTATGTTACTTGCTGTAAAGGAAAATAAACAAC 738

     286 lnValValLysPheLeuIleLysLysLysAlaAsnLeuAsnAlaLeuAsp 302
         :::::|||:::||||||:::||||||||||||:::::::::|||:::|||
     739 ATATTGTGGAATTTTTAGTAAAGAAGAAAGCAAGTATTCACGCAGTTGAT 788

     303 ArgTyrGlyArgThrAlaLeuIleLeuAlaValCysCysGlySerAlaSe 319
         :::   |||
     789 CAGCTTGGA..................................... 797

     319 rIleValSerLeuLeuLeuGluGlnAsnIleAspValSerSerGlnAspL 336

     797 ..................................................... 797

     336 euSerGlyGlnThrAlaArgGluTyrAlaValSerSerHisHisHisVal 352
           |||::::::      ||||||
     798 ..AGCAACAGGCAAATGTTTGAATAT..................... 821

     353 IleCysGlnLeuLeuSerAspTyrLysGluLysGlnMetLeuLysIleSe 369
                        |||   |||:::   |||
     822 ..................GATGGGAAGAGACTACAA.............AG 841

     369 rSerGluAsnSerAsnPro..............GluAsnValSerArgT 381
          ||||||||||||||||||              |||:::   |||   |
     842 ATCTGAAAACAGCAATCCAGTGGATAATGGTTCTGAAGATGGCTCTTTAA 891
```

```
     381 hrArgAsn 383
         |||||:::
     892 CCAGGTCT 899

seq_name: gb_htc:AK015948

seq_documentation_block:
LOCUS       AK015948               1054 bp    mRNA    linear   HTC 19-JAN-2002
DEFINITION  Mus musculus adult male testis cDNA, RIKEN full-length enriched
            library, clone:4930532L20:homolog to WUGSC:H_DJ1035002.1 PROTEIN
            (FRAGMENT), full insert sequence.
ACCESSION   AK015948
VERSION     AK015948.1  GI:12854490
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) adult male testis cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:4930532L20.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2  (sites)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3  (sites)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4  (sites)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 1054)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C.,
            Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T.,
            Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,F.,
            Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T.,
            Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S.,
            Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K.,
            Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J.,
            Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,
            Schriml,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamanaka,I.,
            Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and
            Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. cDNA went
            through one round of normalization to Rot = 10.0 and subtraction to
            Rot = 100.0. Second strand cDNA was prepared with the primer
            adapter of sequence [5'
            GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cleaved
            with BamHI and XhoI. Vector: a modified pBluescript KS(+) after
            bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3'
            end: BamHI. Host: DH10B.
FEATURES              Location/Qualifiers
     source           1. .1054
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="MGD:MGI:1908511"
                      /db_xref="taxon:10090"
                      /clone="4930532L20"
                      /sex="male"
                      /tissue_type="testis"
                      /clone_lib="RIKEN full-length enriched mouse cDNA library"
                      /dev_stage="adult"
     CDS              110. .949
                      /note="data source:SPTR, source key:Q9UDM3, evidence:ISS
                      homolog to WUGSC:H_DJ1035002.1 PROTEIN (FRAGMENT)
                      putative"
                      /codon_start=1
                      /protein_id="BAB30047.1"
                      /db_xref="GI:12854491"
                      /translation="MKKFFPFRGKRKTDDSHSHSSEVPISLAKTAPPNLSIGGGYRLR
                      DKHLKKLHKAATIGNEQKLKDYLERKKYNVNGRDKRSRTPLHLACANGYTNIVSLLIE
                      NQCKINVQDSENRTPLIKAVECQQESCATVLLLHGADPNLVDVYSNTALHYAVCGQNI
                      SLANKLLQYKANLEAKNKDGHTPLLLAVAENNNENMVKFLLKKGADVNASDKNHRTAIM
                      IALIVEPTSSVKLLLQQDTDLAHKDIYGFTAEEYASFNGFTMYHHITANNENKKKTEQ
                      TAY"
     polyA_signal     1033. .1038
                      /note="putative"
     polyA_site       1054
                      /note="putative"
BASE COUNT      346 a    209 c    207 g    290 t      2 others
ORIGIN


alignment_scores:
            Quality:  517.50             Length:    338
              Ratio:  2.453                Gaps:      8
 Percent Similarity:  62.426   Percent Identity: 36.982

alignment_block:
 US-09-810-936-304 x AK015948   ..

 Align seg 1/1  to: AK015948  from: 1  to: 1054

      72 CysArgGlySerGlyLysSerAsnValGlyAlaSerGlyAspHisAspAs 88
         ||||||:::   |||:::   :::   |||:::   |||
      14 TGCAGATCGAGAGGCAATTGTGAACAAGGTGGTTGCGGCAGAGGACCCGG 63

      88 pSerAlaMetLysThrLeuArgAsn.................LysMetG 99
               ::::::      ||||||                     |||
      64 GCTGTTGGTTAGGATTCCCAGGAACATCTTACTTAGAGATCTTACCATGA 113

      99 lyLysTrpCysCysHisCysPheProCysCysArgGlySerGlyLysSer 115
           |||:::          ||||||     ||||||:::   |||:::
     114 AGAAGTTC............TTCCCC...TTCAGGGGAAAAAGGAAGACA 148
```

```
    116 LysValGlyAlaTrpGlyAspTyrAspAspSer..................  126
                                 ||||||||||
    149 .........................GATGATTCCCACAGCCACTCCTCAGA  174

    127 ................AlaPheMetGluPro.....................  131
                        |||          |||
    175 AGTGCCCATTAGCCTGGCCAAAACCGCACCGCCCAACCTGTCCATAGGTG  224

    132 ..ArgTyrHisValArgGlyGluAspLeuAspLysLeuHisArgAlaAla  147
          |||::::::|||   :::   |||   |||||||||:::||||||
    225 GAGGATATCGCCTTCGAGATAAGCACTTAAAGAAACTTCATAAAGCTGCT  274

    148 TrpTrpGlyLysValProArgLysAspLeuIleValMetLeuArgAspTh  164
          |||:::     :::          :::       :::
    275 ACCATTGGCAATGAGCAGAAGCTGAAGGATTACCTTGAACGCAAGAAATA  324

    164 rAspValAsnLysLysAspLysGlnLysArgThrAlaLeuHisLeuAlaS  181
         :::||||||   :::|||||:::::||||||   ||||||||||||
    325 CAATGTGAATGGGCGGGACAAAAGAAGCAGAACACCTTTGCATTTGGCCT  374

    181 erAlaAsnGlyAsnSerGluValValLysLeuLeuLeuAspArgArgCys  197
          |||||||||   ::::::::::|||:::||||||::::::::::::|||
    375 GTGCTAATGGATATACAAATATTGTCTCTCTCTTAATTGAGAATCAATGC  424

    198 GlnLeuAsnValLeuAspAsnLysLysArgThrAlaLeuIleLysAlaVa  214
         ::::::||||||   |||::::::::::||||||   ||||||||||||||
    425 AAAATTAATGTCCAGGATAGTGAAAACAGAACCCCATTGATTAAGGCAGT  474

    214 lGlnCysGlnGluAspGluCysAlaLeuMetLeuLeuGluHisGlyThrA  231
        |:::||||||::::::::::||||||   :::||||||   ||||||:::|
    475 AGAGTGTCAACAGGAGAGTTGTGCTACGGTTCTTCTTCTTCATGGTGCAG  524

    231 spProAsnIleProAspGluTyrGlyAsnThrThrLeuHisTyrAlaIle  247
        ||||||||:::   |||   |||:::||||||:::||||||||||||:::
    525 ATCCAAATCTGGTAGATGTTTATAGTAATACGGCTCTCCATTATGCTGTT  574

    248 TyrAsnGluAspLysLeuMetAlaLysAlaLeuLeuLeuTyrGlyAlaAs  264
         :::::::::       :::|||:::   ||||||   |||   |||::
    575 TGTGGCCAAAATATTTCATTAGCCAATAAATTGCTTCAATATAAAGCCAA  624

    264 pIleGluSerLysAsnLysHisGlyLeuThrProLeuLeuLeuGlyValH  281
         ::::|||:::|||||||||   |||   ||||||||||||||||:::|||
    625 TCTTGAAGCCAAAAATAAGGACGGTCACACTCCGCTTTTACTTGCTGTTG  674

    281 isGluGlnLysGlnGlnValValLysPheLeuIleLysLysLysAlaAsn  297
         |||:::::::::::::::|||||||||||||:::|||||||   |||:::
    675 CTGAAAATAATGAAAATATGGTAAAATTCCTTCTGAAGAAAGGAGCAGAT  724

    298 LeuAsnAlaLeuAspArgTyrGlyArgThrAlaLeuIleLeuAlaValCy  314
         :::|||||||   |||:::      |||||||||:::::::::|||:::
    725 GTAAATGCATCAGATAAAAACCACAGAACAGCCATCATGATTGCTCTGAT  774

    314 sCysGlySerAlaSerIleValSerLeuLeuLeuGluGlnAsnIleAspV  331
                 :::|||   |||:::||||||||||:::|||:::   |||:
    775 TGTTGAACCAACAAGTTCTGTAAAACTTCTTCTTCAGCAAGATACGGACC  824

    331 alSerSerGlnAspLeuSerGlyGlnThrAlaArgGluTyrAla......  345
         :::::   :::|||:::   |||   ||||||:::|||||||||
    825 TAGCCCACAAAGATATTTATGGATTTACAGCTGAGGAATATGCTTCATTT  874

    346 ......ValSerSerHisHisHisValIleCysGlnLeuLeuSerAspTy  360
                :::   :::||||||
    875 AATGGCTTTACTATGTATCACCAT.........................  898

    360 rLysGluLysGlnMetLeuLysIleSerSerGluAsnSerAsnProGluA  377
                           |||:::::::::|||:::|||   ::::
    899 .....................ATAACTGCCAATAATGAGAACAAGAAGA  926

    377 snValSerArgThr 381
         ::::::::::|||
    927 AAACAGAGCAAACA 940

seq_name: gb_est2:BG720647
```

```
seq_documentation_block:
LOCUS       BG720647               694 bp    mRNA    linear   EST 08-MAY-2001
DEFINITION  602692528F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4824630 5',
            mRNA sequence.
ACCESSION   BG720647
VERSION     BG720647.1  GI:13999834
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 694)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
             cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10736  row: h  column: 07
            High quality sequence stop: 694.
FEATURES             Location/Qualifiers
     source          1. .694
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4824630"
                     /clone_lib="NIH_MGC_97"
                     /lab_host="DH10B"
                     /note="Organ: testis; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
                     ); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTVN-3',
                     size-selected for  average insert size 2.2 kb and
                     normalized to ROT 5. This is a primary library enriched
                     for full-length clones and constructed using the
                     Cap-trapper method (Carninci, in preparation).  Library
                     constructed by M. Brownstein (NIMH/NHGRI, National
                     Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT      108 a    173 c    228 g    185 t
ORIGIN


alignment_scores:
            Quality:  496.00             Length:    105
              Ratio:  5.167                Gaps:      0
 Percent Similarity:  91.429   Percent Identity: 83.810

alignment_block:
 US-09-810-936-304 x BG720647   ..

 Align seg 1/1  to: BG720647  from: 1  to: 694

       1 MetValValGluValAspSerMetProAlaAlaSerSerValLysLysPr 17
         ||||||:::|||||||   |||||||||||||||||||:::|||||||||||
     376 ATGGTGGCTGAGGTTTGTTCAATGCCCGCTGCCTCTGCTGTGAAGAAGCC 425

      17 oPheGlyLeuArgSerLysMetGlyLysTrpCysCysArgCysPheProC 34
         ||||   |||||||||||||||||||||||||||   :::   |||||||
     426 ATTTGATCTCAGGAGCAAGATGGGCAAGTGGTGCCACCACCGCTTCCCCT 475

      34 ysCysArgGluSerGlyLysSerAsnValGlyThrSerGlyAspHisAsp 50
         ||||||||   |||||||||||||||:::|||||||||||||||||||||
     476 GCTGCAGGGGGAGCGGCAAGAGCAACATGGGCACTTCTGGAGACCACGAC 525

      51 AspSerAlaMetLysThrLeuArgSerLysMetGlyLysTrpCysArgHi 67
         ||||||   |||||||||||||||||||||||||||||   |||:::||
     526 GACTCCTTTATGAAGACGCTCAGGAGCAAGATGGGCAAGTGTTGCCACCA 575
```

```
     67 sCysPheProCysCysArgGlySerGlyLysSerAsnValGlyAlaSerG 84
        ||||||||||||||||||||||||||   ||||||||||||:::||||
    576 CTGCTTCCCCTGCTGCAGGGGGAGCGGCACGAGCAATGTGGGCACTTCTG 625

     84 lyAspHisAspAspSerAlaMetLysThrLeuArgAsnLysMetGlyLys 100
        ||||||||||:::|||   ||||||||||||||:::||||||||||||
    626 GAGACCATGACAACTCCTTTATGAAGACACTCAGGAGCAAGATGGGCAAG 675

    101 TrpCysCysHisCys 105
        ||||||||||||||
    676 TGGTGCTGTCACTGT 690

seq_name: gb_htc:AK015165

seq_documentation_block:
LOCUS       AK015165              1400 bp    mRNA    linear   HTC 19-JAN-2002
DEFINITION  Mus musculus adult male testis cDNA, RIKEN full-length enriched
            library, clone:4930418I21:similar to ANKYRIN-LIKE PROTEIN, full
            insert sequence.
ACCESSION   AK015165
VERSION     AK015165.1  GI:12853397
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) adult male testis cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:4930418I21.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2  (sites)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3  (sites)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4  (sites)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 1400)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C.,
            Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T.,
            Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,F.,
            Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T.,
            Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S.,
            Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K.,
            Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J.,
            Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,
            Schriml,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamanaka,I.,
            Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and
            Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Gemomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. cDNA went
            through one round of normalization to Rot = 10.0 and subtraction to
            Rot = 100.0. Second strand cDNA was prepared with the primer
            adapter of sequence [5'
            GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved
            with BamHI and XhoI. Vector: a modified pBluescript KS(+) after
            bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3'
            end: BamHI. Host: DH10B.
FEATURES             Location/Qualifiers
     source          1. .1400
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="MGD:MGI:1898762"
                     /db_xref="taxon:10090"
                     /clone="4930418I21"
                     /sex="male"
                     /tissue_type="testis"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     CDS             162. .1193
                     /note="data source:SPTR, source key:Q9EQ29, evidence:ISS
                     putative
                     similar to ANKYRIN-LIKE PROTEIN"
                     /codon_start=1
                     /protein_id="BAB29733.1"
                     /db_xref="GI:12853398"
                     /translation="MKKIFGAKEKTPLGFCDAPQMTVSEFFFGENPNYPKYHTTYRPL
                     GHIHRVAAEGDAARMEILLTLGQCNVYHRDRKDRTALHFACVYGRLPVVTILVKNNCE
                     IDALDKNHTTPLMKSVQCWKQKCATVLLEHGADPNIRDSSGNSALHYAVYNGHQEMAS
                     LLLQYNADIEQKTKDGFTPLLLALREKRVEVAEFLVRMGADIHVVDELQRNTLIYAIR
                     CGSKDLSVLLLERGIDFFYKDVFGWTALRYAIEGHCTFRQTLLDFEESLHSNKKDKEP
                     ELQVEPSTSCILAKQIDAGNDSLARISSCPSPETPVLTMKEDENSDARVNCSSVELLP
                     WPKFVFGLF"
BASE COUNT      409 a    293 c    353 g    345 t
ORIGIN


alignment_scores:
           Quality:  493.50              Length:    249
             Ratio:  2.668                 Gaps:      3
 Percent Similarity:  74.297   Percent Identity: 40.964

alignment_block:
 US-09-810-936-304 x AK015165   ..

 Align seg 1/1  to: AK015165  from: 1  to: 1400

    131 ProArgTyrHisValArgGlyGluAspLeuAspLysLeuHisArgAlaAl 147
        |||:::||||||:::      :::   |||      :::||||||::||
    264 CCGAAATACCATACAACATACAGACCGTTGGGGCATATTCACAGGGTTGC 313

    147 aTrpTrpGlyLysValProArgLysAspLeuIleValMetLeuArgAspT 164
        |      |||   :::   |||   ::::::::::::   |||   :::
```

```
    314 TGCTGAGGGTGATGCAGCCCGGATGGAGATTCTCCTCACGCTTGGACAAT 363

    164 hrAspValAsnLysLysAspLysGlnLysArgThrAlaLeuHisLeuAla 180
    364 GCAACGTGTACCACAGAGACCGCAAGGACAGGACAGCTTTGCACTTTGCC 413

    181 SerAlaAsnGlyAsnSerGluValValLysLeuLeuLeuAspArgArgCy 197
    414 TGTGTCTATGGCCGGCTACCAGTAGTGACCATTCTAGTGAAGAATAATTG 463

    197 sGlnLeuAsnValLeuAspAsnLysLysArgThrAlaLeuIleLysAlaV 214
    464 TGAGATCGATGCCTTGGACAAGAACCATACCACACCTCTGATGAAGTCTG 513

    214 alGlnCysGlnGluAspGluCysAlaLeuMetLeuLeuGluHisGlyThr 230
    514 TCCAATGCTGGAAGCAGAAATGTGCGACCGTGCTGCTGGAGCACGGCGCT 563

    231 AspProAsnIleProAspGluTyrGlyAsnThrThrLeuHisTyrAlaIl 247
    564 GATCCGAACATTAGAGACAGCAGTGGCAACAGTGCCTTGCACTACGCTGT 613

    247 eTyrAsnGluAspLysLeuMetAlaLysAlaLeuLeuLeuTyrGlyAlaA 264
    614 GTACAACGGGCATCAGGAGATGGCCTCCCTACTGCTTCAGTACAACGCTG 663

    264 spIleGluSerLysAsnLysHisGlyLeuThrProLeuLeuLeuGlyVal 280
    664 ATATCGAACAAAAAACCAAGGATGGCTTCACACCACTGTTGCTGGCACTC 713

    281 HisGluGlnLysGlnGlnValValLysPheLeuIleLysLysLysAlaAs 297
    714 CGAGAAAAGCGGGTGGAGGTGGCAGAGTTTTTAGTAAGAATGGGAGCAGA 763

    297 nLeuAsnAlaLeuAspArgTyrGlyArgThrAlaLeuIleLeuAlaValC 314
    764 CATACACGTGGTAGATGAGCTGCAAAGAAACACCTTGATATACGCTATAA 813

    314 ysCysGlySerAlaSerIleValSerLeuLeuLeuGluGlnAsnIleAsp 330
    814 GATGTGGATCAAAAGATCTGTCTGTATTACTCCTTGAGAGAGGTATTGAC 863

    331 ValSerSerGlnAspLeuSerGlyGlnThrAlaArgGluTyrAlaValSe 347
    864 TTCTTCTATAAGGATGTCTTCGGATGGACAGCATTACGCTACGCTATTGA 913

    347 rSerHisHisHisValIleCys.........GlnLeuLeuSerAspTyrL 361
    914 AGGCCAC............TGTACTTTTAGACAGACACTTCTGGATTTTG 951

    361 ysGluLysGlnMetLeuLysIleSerSerGluAsnSerAsnProGlu 376
    952 AAGAAAGT......CTACATAGTAACAAAAAAGATAAGGAACCAGAA 992

seq_name: gb_htc:AK015473

seq_documentation_block:
LOCUS       AK015473               1441 bp    mRNA    linear   HTC 19-JAN-2002
DEFINITION  Mus musculus adult male testis cDNA, RIKEN full-length enriched
            library, clone:4930456F22:similar to ANKYRIN-LIKE PROTEIN, full
            insert sequence.
ACCESSION   AK015473
VERSION     AK015473.1  GI:12853831
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) adult male testis cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:4930456F22.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2  (sites)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3  (sites)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4  (sites)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 1441)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C.,
            Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T.,
            Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,F.,
            Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T.,
            Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S.,
            Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K.,
            Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J.,
            Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,
            Schriml,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamanaka,I.,
            Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and
            Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Gemomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. cDNA went
            through one round of normalization to Rot = 10.0 and subtraction to
            Rot = 100.0. Second strand cDNA was prepared with the primer
            adapter of sequence [5'
            GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved
            with BamHI and XhoI. Vector: a modified pBluescript KS(+) after
            bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3'
            end: BamHI. Host: DH10B.
FEATURES              Location/Qualifiers
     source           1. .1441
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
```

```
                     /db_xref="MGD:MGI:1897164"
                     /db_xref="taxon:10090"
                     /clone="4930456F22"
                     /sex="male"
                     /tissue_type="testis"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     CDS             162. .1193
                     /note="data source:SPTR, source key:Q9EQ29, evidence:ISS
                     putative
                     similar to ANKYRIN-LIKE PROTEIN"
                     /codon_start=1
                     /protein_id="BAB29861.1"
                     /db_xref="GI:12853832"
                     /translation="MKKIFGAKEKTPLGFCDAPQMTVSEFFFGENPNYPKYHTTYRPL
                     GHIHRVAAEGDAARMEILLTLGQCNVYHRDRKDRTALHFACVYGRLPVVTILVKNNCE
                     IDALDKNHTTPLMKSVQCWKQKCATVLLEHGADPNIRDSSGNSALHYAVYNGHQEMAS
                     LLLQYNADIEQKTKDGFTPLLLALREKRVEVAEFLVRMGADIHVVDELQRNTLIYAIR
                     CGSKDLSVLLLEKGIDFFYKDVFGWTALRYAIEGHCTFRQTLLDFEESLHSNKKDKEP
                     ELQVEPSTSCILAKQIDAGNDSLARISSCPSPETPVLTMKEDENSDARVNCSSVELLP
                     WPKFVFGLF"
BASE COUNT      419 a    297 c    361 g    364 t
ORIGIN


alignment_scores:
            Quality:  493.50              Length:    249
              Ratio:  2.668                 Gaps:      3
 Percent Similarity:  74.297   Percent Identity: 40.964

alignment_block:
 US-09-810-936-304 x AK015473   ..

 Align seg 1/1  to: AK015473  from: 1  to: 1441

     131 ProArgTyrHisValArgGlyGluAspLeuAspLysLeuHisArgAlaAl 147
         |||:::||||||:::      :::   |||      :::||||||:::||
     264 CCGAAATACCATACAACATACAGACCGTTGGGGCATATTCACAGGGTTGC 313

     147 aTrpTrpGlyLysValProArgLysAspLeuIleValMetLeuArgAspT 164
         |     |||   :::   |||   ::::::::::   |||   :::
     314 TGCTGAGGGTGATGCAGCCCGGATGGAGATTCTCCTCACGCTTGGACAAT 363

     164 hrAspValAsnLysLysAspLysGlnLysArgThrAlaLeuHisLeuAla 180
         :::|||      :::|||::::::   |||||||||||||||:::|||
     364 GCAACGTGTACCACAGAGACCGCAAGGACAGGACAGCTTTGCACTTTGCC 413

     181 SerAlaAsnGlyAsnSerGluValValLysLeuLeuLeuAspArgArgCy 197
            :::   |||:::      ||||||   :::|||:::   ::::::||
     414 TGTGTCTATGGCCGGCTACCAGTAGTGACCATTCTAGTGAAGAATAATTG 463

     197 sGlnLeuAsnValLeuAspAsnLysLysArgThrAlaLeuIleLysAlaV 214
         |::::::::::::||||||::::::      |||   |||:::|||:::|
     464 TGAGATCGATGCCTTGGACAAGAACCATACCACACCTCTGATGAAGTCTG 513

     214 alGlnCysGlnGluAspGluCysAlaLeuMetLeuLeuGluHisGlyThr 230
         ||||||||   :::::::::||||||   :::|||||||||||||||:::
     514 TCCAATGCTGGAAGCAGAAATGTGCGACCGTGCTGCTGGAGCACGGCGCT 563

     231 AspProAsnIleProAspGluTyrGlyAsnThrThrLeuHisTyrAlaIl 247
         ||||||||||||   |||:::   ||||||::::::||||||||||||::
     564 GATCCGAACATTAGAGACAGCAGTGGCAACAGTGCCTTGCACTACGCTGT 613

     247 eTyrAsnGluAspLysLeuMetAlaLysAlaLeuLeuLeuTyrGlyAlaA 264
         :||||||      :::   |||||||:::   ||||||   |||:::||||
     614 GTACAACGGGCATCAGGAGATGGCCTCCCTACTGCTTCAGTACAACGCTG 663

     264 spIleGluSerLysAsnLysHisGlyLeuThrProLeuLeuLeuGlyVal 280
         ||||||||:::|||:::|||   |||:::||||||||||||||||:::::
     664 ATATCGAACAAAAAACCAAGGATGGCTTCACACCACTGTTGCTGGCACTC 713

     281 HisGluGlnLysGlnGlnValValLysPheLeuIleLysLysLysAlaAs 297
         :::|||::::::   :::|||::::::||||||::::::      |||::
     714 CGAGAAAAGCGGGTGGAGGTGGCAGAGTTTTTAGTAAGAATGGGAGCAGA 763

     297 nLeuAsnAlaLeuAspArgTyrGlyArgThrAlaLeuIleLeuAlaValC 314
         :::::::::::::|||:::      |||:::::|||||||   |||:::
     764 CATACACGTGGTAGATGAGCTGCAAAGAAACACCTTGATATACGCTATAA 813

     314 ysCysGlySerAlaSerIleValSerLeuLeuLeuGluGlnAsnIleAsp 330
           |||||||||   ::::::      ||||||||||||:::::||||||
     814 GATGTGGATCAAAAGATCTGTCTGTATTACTCCTTGAGAAAGGTATTGAC 863

     331 ValSerSerGlnAspLeuSerGlyGlnThrAlaArgGluTyrAlaValSe 347
                 :::|||:::   |||   ||||||   :::||||||:::::
     864 TTCTTCTATAAGGATGTCTTCGGATGGACAGCATTACGCTACGCTATTGA 913

     347 rSerHisHisHisValIleCys.........GlnLeuLeuSerAspTyrL 361
         :::::|||            |||         |||   |||   |||::::
     914 AGGCCAC............TGTACTTTTAGACAGACACTTCTGGATTTTG 951

     361 ysGluLysGlnMetLeuLysIleSerSerGluAsnSerAsnProGlu 376
         ::|||:::      |||      ::::::::::::::::::||||||
     952 AAGAAAGT......CTACATAGTAACAAAAAAGATAAGGAACCAGAA 992

seq_name: gb_htc:AK015012

seq_documentation_block:
LOCUS       AK015012              1456 bp    mRNA    linear   HTC 19-JAN-2002
DEFINITION  Mus musculus adult male testis cDNA, RIKEN full-length enriched
            library, clone:4921537P18:Ank repeat containing protein, full
            insert sequence.
ACCESSION   AK015012
VERSION     AK015012.1  GI:12853181
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) adult male testis cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:4921537P18.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2  (sites)
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3  (sites)
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4  (sites)
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5  (bases 1 to 1456)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C.,
```

```
            Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T.,
            Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,F.,
            Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T.,
            Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S.,
            Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K.,
            Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J.,
            Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,
            Schriml,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamanaka,I.,
            Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and
            Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Gemomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. Second
            strand cDNA was prepared with the primer adapter of sequence [5'
            GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved
            with BamHI and XhoI. cDNA of size comprised longer than 7 kb was
            selected before cloning. Vector: a modified pBluescript KS(+) after
            bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3'
            end: BamHI. Host: DH10B.
FEATURES              Location/Qualifiers
     source           1. .1456
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="MGD:MGI:1903425"
                      /db_xref="taxon:10090"
                      /clone="4921537P18"
                      /sex="male"
                      /tissue_type="testis"
                      /clone_lib="RIKEN full-length enriched mouse cDNA library"
                      /dev_stage="adult"
     CDS              150. .>1456
                      /note="Ank repeat containing protein
                      data source:Pfam, source key:PF00023, evidence:ISS
                      putative"
                      /codon_start=1
                      /protein_id="BAB29670.1"
                      /db_xref="GI:12853182"
                      /translation="MSFSMKKIFGAKEKTPLGFCDAPQMTVSEFFFGENPNYPKYHTT
                      YRPLGHIHRVAAEGDAARMEILLTLGQCNVYHRDRKDRTALHFACVYGRLPVVTILVK
                      NNCEIDALDKNHTTPLMKSVQCWKQKCATVLLEHGADPNIRDSSGNSALHYAVYNGHQ
                      EMASLLLQYNADIEQKTKDGFTPLLLALREKRVEVAEFLVRMGADIHVVDELQRNTLI
                      YAIRCGSKDLSVLLLEKGIDFFYKDVFGWTALRYAIEGHCTFRQTLLDFEESLHSNKK
                      DKEPELQVEPSTSCILAKQIDAGNDSLARISSCPSPETPVLTMKEDENSDARSNSESK
                      PPNCAGNSPAAIDKAFGNRGKRTVEAWVEKDPSSESAAEVRDSVTNEAVERIKLLPSE
                      TELELMSVEEDVSHESENDHPLRESEHLPKIKVGHSSRSGYQMG"
BASE COUNT      437 a    309 c    372 g    338 t
ORIGIN


alignment_scores:
           Quality:  493.50            Length:    249
             Ratio:  2.668               Gaps:      3
 Percent Similarity:  74.297   Percent Identity: 40.964

alignment_block:
 US-09-810-936-304 x AK015012   ..
```

```
Align seg 1/1  to: AK015012  from: 1  to: 1456

    131 ProArgTyrHisValArgGlyGluAspLeuAspLysLeuHisArgAlaAl 147
        |||:::|||||:::      :::   |||     :::|||||:::||
    264 CCGAAATACCATACAACATACAGACCGTTGGGGCATATTCACAGGGTTGC 313

    147 aTrpTrpGlyLysValProArgLysAspLeuIleValMetLeuArgAspT 164
        |   |||   :::   |||   ::::::::::::   |||   :::
    314 TGCTGAGGGTGATGCAGCCCGGATGGAGATTCTCCTCACGCTTGGACAAT 363

    164 hrAspValAsnLysLysAspLysGlnLysArgThrAlaLeuHisLeuAla 180
          :::|||     :::|||:::::   |||||||||||||||:::|||
    364 GCAACGTGTACCACAGAGACCGCAAGGACAGGACAGCTTTGCACTTTGCC 413

    181 SerAlaAsnGlyAsnSerGluValValLysLeuLeuLeuAspArgArgCy 197
          :::   |||:::     ||||||   :::|||:::   ::::::||
    414 TGTGTCTATGGCCGGCTACCAGTAGTGACCATTCTAGTGAAGAATAATTG 463

    197 sGlnLeuAsnValLeuAspAsnLysLysArgThrAlaLeuIleLysAlaV 214
        |:::::::::::|||||:::::      |||   |||:::|||:::|
    464 TGAGATCGATGCCTTGGACAAGAACCATACCACACCTCTGATGAAGTCTG 513

    214 alGlnCysGlnGluAspGluCysAlaLeuMetLeuLeuGluHisGlyThr 230
        ||||||||   :::::::::||||||   :::|||||||||||||||:::
    514 TCCAATGCTGGAAGCAGAAATGTGCGACCGTGCTGCTGGAGCACGGCGCT 563

    231 AspProAsnIleProAspGluTyrGlyAsnThrThrLeuHisTyrAlaIl 247
        ||||||||||||   |||:::   ||||||:::::|||||||||||||:
    564 GATCCGAACATTAGAGACAGCAGTGGCAACAGTGCCTTGCACTACGCTGT 613

    247 eTyrAsnGluAspLysLeuMetAlaLysAlaLeuLeuLeuTyrGlyAlaA 264
        :||||||      :::   ||||||:::   ||||||   |||:::||||
    614 GTACAACGGGCATCAGGAGATGGCCTCCCTACTGCTTCAGTACAACGCTG 663

    264 spIleGluSerLysAsnLysHisGlyLeuThrProLeuLeuLeuGlyVal 280
        ||||||||:::|||:::|||   |||:::|||||||||||||||:::::
    664 ATATCGAACAAAAAACCAAGGATGGCTTCACACCACTGTTGCTGGCACTC 713

    281 HisGluGlnLysGlnGlnValValLysPheLeuIleLysLysLysAlaAs 297
        :::|||:::::::   :::|||:::::::||||||:::::      |||::
    714 CGAGAAAAGCGGGTGGAGGTGGCAGAGTTTTTAGTAAGAATGGGAGCAGA 763

    297 nLeuAsnAlaLeuAspArgTyrGlyArgThrAlaLeuIleLeuAlaValC 314
        :::::::::::::|||:::      |||:::::::||||||   |||:::
    764 CATACACGTGGTAGATGAGCTGCAAAGAAACACCTTGATATACGCTATAA 813

    314 ysCysGlySerAlaSerIleValSerLeuLeuLeuGluGlnAsnIleAsp 330
          |||||||||   ::::::      ||||||||||||:::::||||||
    814 GATGTGGATCAAAAGATCTGTCTGTATTACTCCTTGAGAAAGGTATTGAC 863

    331 ValSerSerGlnAspLeuSerGlyGlnThrAlaArgGluTyrAlaValSe 347
                ::::|||:::   |||   ||||||   :::|||||||:::::
    864 TTCTTCTATAAGGATGTCTTCGGATGGACAGCATTACGCTACGCTATTGA 913

    347 rSerHisHisHisValIleCys.........GlnLeuLeuSerAspTyrL 361
        :::::||           |||         |||   |||   |||::::
    914 AGGCCAC............TGTACTTTAGACAGACACTTCTGGATTTTG 951

    361 ysGluLysGlnMetLeuLysIleSerSerGluAsnSerAsnProGlu 376
        ::|||:::      |||      ::::::::::::::::::||||||
    952 AAGAAAGT......CTACATAGTAACAAAAAAGATAAGGAACCAGAA 992

seq_name: gb_est2:BI830348

seq_documentation_block:
LOCUS       BI830348               856 bp    mRNA    linear   EST 04-OCT-2001
DEFINITION  603073084F1 NIH_MGC_119 Homo sapiens cDNA clone IMAGE:5164868 5',
            mRNA sequence.
ACCESSION   BI830348
VERSION     BI830348.1  GI:15941898
KEYWORDS    EST.
SOURCE      human.
```

```
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 856)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11409  row: h  column: 21
            High quality sequence stop: 842.
FEATURES             Location/Qualifiers
     source          1. .856
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5164868"
                     /clone_lib="NIH_MGC_119"
                     /tissue_type="medulla"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: EcoRV (destroyed); RNA source normal medulla from
                     anonymous male age 27.  Library  is oligo-dT primed and
                     directionally cloned (EcoRV site is destroyed  upon
                     cloning).  Average insert size 1.3 kb, insert size range
                     0.9-3 kb.  Library is normalized and enriched for
                     full-length clones and  was constructed by C. Gruber
                     (Invitrogen).  Research Genetics  tracking code 013. Note:
                     this is a NIH_MGC Library."
BASE COUNT      266 a    160 c    206 g    224 t
ORIGIN


alignment_scores:
            Quality:  475.00             Length:    219
              Ratio:  2.844                Gaps:      1
 Percent Similarity:  76.256   Percent Identity: 45.205

alignment_block:
 US-09-810-936-304 x BI830348   ..

 Align seg 1/1  to: BI830348  from: 1  to: 856

     130 GluProArgTyrHisValArgGlyGluAspLeuAspLysLeuHisArgAl 146
         :::|||   |||::::::|||   :::|||||   |||||||||||||
     146 CAGCCAGGCTACAACCTTCGAGAAAAGGATTTAAAGAAACTTCACAGAGC 195

     146 aAlaTrpTrpGlyLysValProArgLysAspLeuIleValMetLeuArgA 163
         ||||     |||  :::  :::          ::::::::::::
     196 TGCTTCAGTCGGGGATTTGAAGAAGCTGAAGGAATACCTTCAGATCAAGA 245

     163 spThrAspValAsnLysLysAspLysGlnLysArgThrAlaLeuHisLeu 179
           |||||||||   :::||||||:::   ||||||   |||||||||
     246 AATATGATGTAAATATGCAGGACAAAAAATACAGAACACCTTTGCACCTA 295

     180 AlaSerAlaAsnGlyAsnSerGluValValLysLeuLeuLeuAspArgAr 196
         |||   |||||||||:::::::::||||||   :::|||:::::::::::
     296 GCCTGTGCTAATGGACATACAGATGTTGTACTTTTCCTAATTGAGCAACA 345

     196 gCysGlnLeuAsnValLeuAspAsnLysLysArgThrAlaLeuIleLysA 213
         :|||:::::||||||   |||:::::::::::::::   ||||||||||
     346 ATGCAAAATAAATGTCCGGGATAGTGAAAACAAATCCCCATTGATTAAGG 395

     213 laValGlnCysGlnGluAspGluCysAlaLeuMetLeuLeuGluHisGly 229
         ||||||||||||||:::::::::::||||||   :::|||||||::   |||
     396 CAGTACAGTGTCAAAATGAGGATTGTGCTACTATTCTTCTAAACTTTGGT 445

     230 ThrAspProAsnIleProAspGluTyrGlyAsnThrThrLeuHisTyrAl 246
         :::||||||::::::   |||         ||||||:::||||||||||
     446 GCAGACCCAGATCTGAGGGATATTCGTTATAATACTGTTCTTCACTATGC 495

     246 aIleTyrAsnGluAspLysLeuMetAlaLysAlaLeuLeuLeuTyrGlyA 263
         |:::   :::::::::      :::::::::   ||||||   |||   |
     496 TGTTTGTGGTCAAAGTTTGTCATTAGTTGAAAAACTGCTTGAATACGAAG 545

     263 laAspIleGluSerLysAsnLysHisGlyLeuThrProLeuLeuLeuGly 279
         ||||||:::|||:::||||||||||   |||   ||||||||||||::::::
     546 CTGATCTTGAAGCGAAAAATAAGGATGGGTATACTCCACTATTAGTTGCC 595

     280 ValHisGluGlnLysGlnGlnValValLysPheLeuIleLysLysLysAl 296
         |||   :::::::::   ::::::||||||||||||||:::::::|||   ||
     596 GTTATTAACAATAATCCAAAAATGGTAAAATTTCTTCTGGAGAAAGGGGC 645

     296 aAsnLeuAsnAlaLeuAspArgTyrGlyArgThrAlaLeuIleLeuAlaV 313
         |:::::||||||   |||:::|||   ||||||||||||||||||||||
     646 TGATGTGAATGCTTCAGATAATTATCAAAGAACAGCCCTTATTCTTGCTG 695

     313 alCysCysGlySerAlaSerIleValSerLeuLeuLeuGluGlnAsnIle 329
         ||                  :::|||:::||||||||:::|||::::::
     696 TCAGTGGTGAACCACCATGTTTAGTAAAGCTTCTTCTTCAGCAAGGTGTG 745

     330 ..AspValSerSerGlnAspLeuSerGlyGlnThrAlaArgGluTyrAla 345
           ::::::    :::       |||   ||||||:::|||||:::
     746 GGAATTTATGTTACGAAGGGTATTGTGGATTCACAGCTGAGGAATATGTT 795

     346 ValSer 347
         :::|||
     796 ATTTCC 801

seq_name: gb_est2:BI827177

seq_documentation_block:
LOCUS       BI827177                828 bp    mRNA    linear   EST 04-OCT-2001
DEFINITION  603077385F1 NIH_MGC_119 Homo sapiens cDNA clone IMAGE:5169141 5',
            mRNA sequence.
ACCESSION   BI827177
VERSION     BI827177.1  GI:15938727
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 828)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11420  row: j  column: 22
            High quality sequence stop: 816.
FEATURES             Location/Qualifiers
     source          1. .828
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5169141"
                     /clone_lib="NIH_MGC_119"
                     /tissue_type="medulla"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: EcoRV (destroyed); RNA source normal medulla from
                     anonymous male age 27.  Library  is oligo-dT primed and
                     directionally cloned (EcoRV site is destroyed  upon
                     cloning).  Average insert size 1.3 kb, insert size range
```

0.9-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 013. Note: this is a NIH_MGC Library."

BASE COUNT 259 a 151 c 196 g 222 t
ORIGIN

alignment_scores:
Quality: 471.00 Length: 213
Ratio: 2.907 Gaps: 1
Percent Similarity: 76.056 Percent Identity: 46.009

alignment_block:
US-09-810-936-304 x BI827177 ..

Align seg 1/1 to: BI827177 from: 1 to: 828

```
    133 TyrHisValArgGlyGluAspLeuAspLysLeuHisArgAlaAlaTrpTr 149
        |||:::::::|||   :::|||||||   ||||||||||||||||||
    156 TACAACCTTCGAGAAAAGGATTTAAAGAAACTTCACAGAGCTGCTTCAGT 205

    149 pGlyLysValProArgLysAspLeuIleValMetLeuArgAspThrAspV 166
         |||   :::   :::              ::::::::::::      ||||
    206 CGGGGATTTGAAGAAGCTGAAGGAATACCTTCAGATCAAGAAATATGATG 255

    166 alAsnLysLysAspLysGlnLysArgThrAlaLeuHisLeuAlaSerAla 182
        |||||   :::||||||:::   ||||||   ||||||||||||   |||
    256 TAAATATGCAGGACAAAAAATACAGAACACCTTTGCACCTAGCCTGTGCT 305

    183 AsnGlyAsnSerGluValValLysLeuLeuLeuAspArgArgCysGlnLe 199
        ||||||:::::::::||||||   :::|||:::::::::::|||:::::
    306 AATGGACATACAGATGTTGTACTTTTCCTAATTGAGCAACAATGCAAAAT 355

    199 uAsnValLeuAspAsnLysLysArgThrAlaLeuIleLysAlaValGlnC 216
        :||||||   |||:::::::::::::::   ||||||||||||||||||
    356 AAATGTCCGGGATAGTGAAAACAAATCCCCATTGATTAAGGCAGTACAGT 405

    216 ysGlnGluAspGluCysAlaLeuMetLeuLeuGluHisGlyThrAspPro 232
        |||||::::::::::||||||   :::|||||:::   |||:::|||||||
    406 GTCAAAATGAGGATTGTGCTACTATTCTTCTAAACTTTGGTGCAGACCCA 455

    233 AsnIleProAspGluTyrGlyAsnThrThrLeuHisTyrAlaIleTyrAs 249
        ::::::   |||         ||||||:::|||||||||||||:::   ::
    456 GATCTGAGGGATATTCGTTATAATACTGTTCTTCACTATGCTGTTTGTGG 505

    249 nGluAspLysLeuMetAlaLysAlaLeuLeuLeuTyrGlyAlaAspIleG 266
        :::::::      :::::::::   ||||||   |||   ||||||:::|
    506 TCAAAGTTTGTCATTAGTTGAAAAACTGCTTGAATACGAAGCTGATCTTG 555

    266 luSerLysAsnLysHisGlyLeuThrProLeuLeuLeuGlyValHisGlu 282
        ||:::|||||||||   |||   ||||||||||||:::::::|||   :::
    556 AAGCGAAAAATAAGGATGGGTATACTCCACTATTAGTTGCCGTTATTAAC 605

    283 GlnLysGlnGlnValValLysPheLeuIleLysLysLysAlaAsnLeuAs 299
        ::::::   ::::::||||||||||||:::::|||   |||:::::::||
    606 AATAATCCAAAAATGGTAAAATTTCTTCTGGAGAAAGGGGCTGATGTGAA 655

    299 nAlaLeuAspArgTyrGlyArgThrAlaLeuIleLeuAlaValCysCysG 316
        ||||   |||:::|||   ||||||||||||||||||||||
    656 TGCTTCAGATAATTATCAAAGAACAGCCCTTATTCTTGCTGTCAGTGGTG 705

    316 lySerAlaSerIleValSerLeuLeuLeuGluGlnAsnIleAspValSer 332
                :::|||:::|||||||||:::|||:::::::   :::
    706 AACCACCATGTTTAGTAAAGCTTCTTCTTCAGCAAGGTGTGGCAATTCAT 755

    333 SerGlnAspLeuSer.GlyGlnThrAlaArgGluTyr 344
                ::|||| |||   |||||||:::|||||
    756 GTTACGAAGGTATCGTGGATTCACAGCTGAGGAATAT 792
```

seq_name: gb_htc:AK005925

seq_documentation_block:

LOCUS AK005925 1209 bp mRNA linear HTC 19-JAN-2002
DEFINITION Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:1700012M14:ANKYRIN-LIKE PROTEIN, full insert sequence.
ACCESSION AK005925
VERSION AK005925.1 GI:12838759
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) adult male testis cDNA to mRNA, clone_lib:RIKEN full-length enriched mouse cDNA library clone:1700012M14.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (sites)
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2 (sites)
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3 (sites)
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4 (sites)
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
REFERENCE 5 (bases 1 to 1209)
AUTHORS Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A., Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,F., Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Schriml,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamanaka,I., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Gemomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN.

```
          Division of Experimental Animal Research in Riken contributed to
          prepare mouse tissues. First strand cDNA was primed with a primer
          [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
          prepared by using trehalose thermo-activated reverse transcriptase
          and subsequently enriched for full-length by cap-trapper. Second
          strand cDNA was prepared with the primer adapter of sequence[5'
          GAGAGAGAGAGCGGCCGCAATTAATTCTCGAGTTAATTAAATTAATCCCCCCCCCCC 3']. cDNA
          was cleaved with XhoI and SstI. Cloning sites, 5' end: XhoI; 3'
          end: SstI. Host: SOLR.
FEATURES             Location/Qualifiers
     source          1. .1209
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="MGD:MGI:1909919"
                     /db_xref="taxon:10090"
                     /clone="1700012M14"
                     /sex="male"
                     /tissue_type="testis"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     CDS             218. .1120
                     /note="ANKYRIN-LIKE PROTEIN
                     data source:SPTR, source key:Q9EQ29, evidence:ISS
                     putative"
                     /codon_start=1
                     /protein_id="BAB24320.1"
                     /db_xref="GI:12838760"
                     /translation="MKKIFKGHSPWGFLSFRHAENQEPPRYLTGYMPVKRIHKAASVG
                     DIPQVQKMLEFGDIDVNVTDRKKRTALHYACAHGQSEMVSLLLWYDCNIEARDREEST
                     ALIKATQRQHEICVKILLENGADSNAVDIHQNTSLHYTVYNKDTTIAAKLLAFNADTE
                     VKTKNGYTPLILAVLENKQEMVELLLQAAANINALDNCKRSALIHAVRTQSKNMISLL
                     LQQGANASLVDIYGATAQSYAVFETFQVLSQGPGPRPELTTKEARHSFDTKDGLVPCS
                     PRFKEDNKRMEENTDAVDELGRYN"
BASE COUNT      332 a    288 c    328 g    261 t
ORIGIN


alignment_scores:
            Quality:  471.00             Length:    225
              Ratio:  2.820                Gaps:      0
 Percent Similarity:  74.222   Percent Identity: 44.889

alignment_block:
 US-09-810-936-304 x AK005925   ..

 Align seg 1/1  to: AK005925  from: 1  to: 1209

     131 ProArgTyrHisValArgGlyGluAspLeuAspLysLeuHisArgAlaAl 147
         |||||||||   :::            :::   :::::::|||:::|||||
     290 CCCAGGTACCTCACTGGGTACATGCCTGTGAAAAGGATCCACAAAGCTGC 339

     147 aTrpTrpGlyLysValProArgLysAspLeuIleValMetLeuArgAspT 164
         |      |||   :::|||:::   :::   ::::::   :::   |||
     340 AAGCGTGGGGGACATACCCCAAGTGCAGAAGATGCTCGAATTTGGGGATA 389

     164 hrAspValAsnLysLysAspLysGlnLysArgThrAlaLeuHisLeuAla 180
           |||||||||      |||:::::|||||||||||||||||||   |||
     390 TTGACGTGAACGTTACAGACCGGAAGAAAAGGACTGCTCTTCACTACGCC 439

     181 SerAlaAsnGlyAsnSerGluValValLysLeuLeuLeuAspArgArgCy 197
           |||:::|||:::||||||:::|||:::|||||||||          ||
     440 TGTGCTCACGGCCAGTCAGAAATGGTGAGCCTCTTGCTGTGGTATGACTG 489

     197 sGlnLeuAsnValLeuAspAsnLysLysArgThrAlaLeuIleLysAlaV 214
         |:::::::::::   |||:::::::::   |||||||||||||||||:
     490 CAACATCGAAGCCCGAGACCGAGAGGAAAGCACAGCTCTGATCAAGGCCA 539

     214 alGlnCysGlnGluAspGluCysAlaLeuMetLeuLeuGluHisGlyThr 230
         ::|||   |||::::::   |||:::   :::||||||||||:::|||::
     540 CACAGCGCCAGCACGAGATCTGTGTGAAGATTCTTCTAGAGAACGGCGCC 589

     231 AspProAsnIleProAspGluTyrGlyAsnThrThrLeuHisTyrAlaIl 247
         |||   |||      |||   :::   ||||||:::||||||||||:::::
     590 GACTCCAACGCCGTCGATATCCATCAGAACACCTCCCTCCACTACACGGT 639

     247 eTyrAsnGluAspLysLeuMetAlaLysAlaLeuLeuLeuTyrGlyAlaA 264
         :||||||:::|||      :::|||      ||||||   ::::::||||
     640 CTACAACAAGGACACCACCATAGCCGCCAAGCTTCTGGCATTCAATGCAG 689

     264 spIleGluSerLysAsnLysHisGlyLeuThrProLeuLeuLeuGlyVal 280
         ||   |||   |||:::|||:::|||   |||||||||:::|||:::|||
     690 ACACCGAAGTTAAGACAAAGAACGGCTACACGCCGCTTATACTTGCGGTG 739

     281 HisGluGlnLysGlnGlnValValLysPheLeuIleLysLysLysAlaAs 297
           |||:::||||||:::::::|||:::::::|||:::::      |||||
     740 CTAGAAAACAAACAGGAGATGGTGGAACTCTTACTGCAGGCTGCAGCCAA 789

     297 nLeuAsnAlaLeuAspArgTyrGlyArgThrAlaLeuIleLeuAlaValC 314
         |:::|||||||||||:::      |||:::||||||||   ||||||
     790 CATAAACGCCCTGGATAACTGCAAGAGGTCCGCCCTTATACATGCTGTGA 839

     314 ysCysGlySerAlaSerIleValSerLeuLeuLeuGluGlnAsnIleAsp 330
               |||   :::::::::|||||||||||||:::|||:::   :::
     840 GAACTCAGTCTAAAAACATGATCAGCCTTCTTCTCCAGCAAGGCGCCAAT 889

     331 ValSerSerGlnAspLeuSerGlyGlnThrAlaArgGluTyrAlaValSe 347
         :::|||      |||:::   |||   ||||||::::::|||||||||
     890 GCATCTTTGGTGGACATCTATGGAGCTACTGCACAAAGTTACGCTGTTTT 939

     347 rSerHisHisHisValIleCysGln 355
          :::      :::|||:::   |||
     940 TGAAACCTTTCAAGTGCTTTCCCAA 964

seq_name: gb_est1:AA910780

seq_documentation_block:
LOCUS       AA910780               539 bp    mRNA    linear   EST 10-JUN-1998
DEFINITION  ol48b04.s1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
            IMAGE:1526671 3' similar to TR:Q92527 Q92527 ANKYRIN MOTIF. ;, mRNA
            sequence.
ACCESSION   AA910780
VERSION     AA910780.1  GI:3050070
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 539)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1548   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 235.
FEATURES             Location/Qualifiers
     source          1. .539
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1526671"
                     /clone_lib="Soares_NFL_T_GBC_S1"
                     /lab_host="DH10B"
                     /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Equal amounts of plasmid DNA from three normalized
                     libraries (fetal lung NbHL19W, testis NHT, and B-cell
                     NCI_CGAP_GCB1) were mixed, and ss circles were made in
                     vitro. Following HAP purification, this DNA was used as
                     tracer in a subtractive hybridization reaction. The driver
                     was PCR-amplified cDNAs from pools of 5,000 clones made
                     from the same 3 libraries. The pools consisted of
                     I.M.A.G.E. clones 297480-302087, 682632-687239,
```

```
                    726408-728711, and 729096-731399. Subtraction by Bento
                    Soares and M. Fatima Bonaldo. "
BASE COUNT      156 a    137 c    138 g    108 t
ORIGIN


alignment_scores:
            Quality:  468.00             Length:    160
              Ratio:  3.545                Gaps:      0
 Percent Similarity:  82.500   Percent Identity: 58.125

alignment_block:
 US-09-810-936-304 x AA910780   ..

 Align seg 1/1  to: AA910780  from: 1  to: 539

     121 GlyAspTyrAspAspSerAlaPheMetGluProArgTyrHisValArgGl 137
         |||:::      :::::::|||:::   :::|||   ||||||||||||
      43 GGCGAGCCGGGGGAGGGCGCCTACTCGCAGCCCGGCTACCACGTCCGAGA 92

     137 yGluAspLeuAspLysLeuHisArgAlaAlaTrpTrpGlyLysValProA 154
          :::||||||   |||:::|||:::||||||      |||:::|||   :
      93 CCGAGATCTCGGCAAGATCCACAAAGCTGCCAGCGCGGGTAATGTGGCGA 142

     154 rgLysAspLeuIleValMetLeuArgAspThrAspValAsnLysLysAsp 170
         ::   :::   |||::::::||||||   :::   :::|||   :::|||
     143 AAGTGCAGCAGATCCTTTTGCTCAGGAAGAATGGCTTGAACGATAGAGAC 192

     171 LysGlnLysArgThrAlaLeuHisLeuAlaSerAlaAsnGlyAsnSerGl 187
         |||::::::|||||||||||||||||||||   :::|||||||:::   ||
     193 AAGATGAACAGGACGGCTCTACATTTGGCCTG.TGCAATGGTCATCCAGA 241

     187 uValValLysLeuLeuLeuAspArgArgCysGlnLeuAsnValLeuAspA 204
         |||||||   ||||||:::||||||:::||||||||||||||||   ||||
     242 AGTAGTAACTCTCCTGGTGGACAGAAAATGCCAGCTCAATGTCTGTGACA 291

     204 snLysLysArgThrAlaLeuIleLysAlaValGlnCysGlnGluAspGlu 220
         ||::::::|||||||||||||:::||||||||||||||||||||||:::::
     292 ACGAAAACAGGACAGCTCTGATGAAGGCTGTACAATGCCAGGAAGAGAAA 341

     221 CysAlaLeuMetLeuLeuGluHisGlyThrAspProAsnIleProAspGl 237
         ||||||   :::|||||||||||||||||:::||||||||||:::   |||
     342 TGTGCAACTATTCTGCTAGAACATGGTGCTGATCCAAATCTTGCGGATGT 391

     237 uTyrGlyAsnThrThrLeuHisTyrAlaIleTyrAsnGluAspLysLeuM 254
          :::||||||||||:::|||||||||||||:::||||||||||||      :
     392 CCATGGCAACACTGCTCTTCACTATGCTGTCTATAATGAGGACATATCAG 441

     254 etAlaLysAlaLeuLeuLeuTyrGlyAlaAspIleGluSerLysAsnLys 270
         ::|||      ||||||||||||   |||:::||||||:::::::|||:::
     442 TAGCAACAAAGCTGCTTTTGTATGATGCAAATATTGAAGCACAAAACCAG 491

     271 HisGlyLeuThrProLeuLeuLeuGlyVal 280
                |||||||||||||||||:::|||
     492 GATGACCTCACACCACTTTTACTTGCAGTA 521

seq_name: gb_est2:BG716974

seq_documentation_block:
LOCUS       BG716974                 916 bp    mRNA    linear   EST 08-MAY-2001
DEFINITION  602689005F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4821576 5',
            mRNA sequence.
ACCESSION   BG716974
VERSION     BG716974.1  GI:13996161
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 916)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
             cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10728  row: i  column: 01
            High quality sequence stop: 832.
FEATURES             Location/Qualifiers
     source          1. .916
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4821576"
                     /clone_lib="NIH_MGC_97"
                     /lab_host="DH10B"
                     /note="Organ: testis; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
                     ); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTVN-3',
                     size-selected for  average insert size 2.2 kb and
                     normalized to ROT 5. This is a primary library enriched
                     for full-length clones and constructed using the
                     Cap-trapper method (Carninci, in preparation).  Library
                     constructed by M. Brownstein (NIMH/NHGRI, National
                     Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT      282 a    174 c    218 g    242 t
ORIGIN


alignment_scores:
            Quality:  468.00             Length:    202
              Ratio:  3.000                Gaps:      0
 Percent Similarity:  77.228   Percent Identity: 46.040

alignment_block:
 US-09-810-936-304 x BG716974   ..

 Align seg 1/1  to: BG716974  from: 1  to: 916

     130 GluProArgTyrHisValArgGlyGluAspLeuAspLysLeuHisArgAl 146
         :::|||   |||:::::::|||  :::|||||||   ||||||||||||||
     161 CAGCCAGGCTACAACCTTCGAGAAAAGGATTTAAAGAAACTTCACAGAGC 210

     146 aAlaTrpTrpGlyLysValProArgLysAspLeuIleValMetLeuArgA 163
         ||||      |||  :::  :::             ::::::::::::
     211 TGCTTCAGTCGGGGATTTGAAGAAGCTGAAGGAATACCTTCAGATCAAGA 260

     163 spThrAspValAsnLysLysAspLysGlnLysArgThrAlaLeuHisLeu 179
           |||||||||  :::|||||:::   ||||||   |||||||||
     261 AATATGATGTAAATATGCAGGACAAAAAATACAGAACACCTTTGCACCTA 310

     180 AlaSerAlaAsnGlyAsnSerGluValValLysLeuLeuLeuAspArgAr 196
         |||   |||||||||:::::::::||||||   ::::|||::::::::::
     311 GCCTGTGCTAATGGACATACAGATGTTGTACTTTTCCTAATTGAGCAACA 360

     196 gCysGlnLeuAsnValLeuAspAsnLysLysArgThrAlaLeuIleLysA 213
         :|||:::::::|||||   |||:::::::::::::   |||||||||
     361 ATGCAAAATAAATGTCCGGGATAGTGAAAACAAATCCCCATTGATTAAGG 410

     213 laValGlnCysGlnGluAspGluCysAlaLeuMetLeuLeuGluHisGly 229
         |||||||||||||||:::::::::||||||   :::|||||:::   |||
     411 CAGTACAGTGTCAAAATGAGGATTGTGCTACTATTCTTCTAAACTTTGGT 460

     230 ThrAspProAsnIleProAspGluTyrGlyAsnThrThrLeuHisTyrAl 246
         :::||||||:::::   |||         ||||||:::||||||||||
     461 GCAGACCCAGATCTGAGGGATATTCGTTATAATACTGTTCTTCACTATGC 510

     246 aIleTyrAsnGluAspLysLeuMetAlaLysAlaLeuLeuLeuTyrGlyA 263
         |:::   :::::::::      :::::::::   ||||||   |||   |
     511 TGTTTGTGGTCAAAGTTTGTCATTAGTTGAAAAACTGCTTGAATACGAAG 560
```

```
263 laAspIleGluSerLysAsnLysHisGlyLeuThrProLeuLeuLeuGly 279
    |||||:::|||:::|||||||||   |||   ||||||||||||:::::
561 CTGATCTTGAAGCGAAAAATAAGGATGGGTATACTCCACTATTAGTTGCC 610

280 ValHisGluGlnLysGlnGlnValValLysPheLeuIleLysLysLysAl 296
    |||   :::::::::   ::::::||||||||||||:::::|||   ||
611 GTTATTAACAATAATCCAAAAATGGTAAAATTTCTTCTGGAGAAAGGGGC 660

296 aAsnLeuAsnAlaLeuAspArgTyrGlyArgThrAlaLeuIleLeuAlaV 313
    |::::::||||||   |||:::|||   ||||||||||||||||||||
661 TGATGTGAATGCTTCAGATAATTATCAAAGAACAGCCCTTATTCTTGCTG 710

313 alCysCysGlySerAlaSerIleVal*SerLeuLeuLeuGluGlnAsnIle* 329
    ||                :::|||:::||||||||||:::   ::::::
711 *TCAGTGGTGAACCACCATGTTTAGTAAAGCTTCTTCTTTCAGCAAGTGTG* 760

330 AspVal 331
    |||:::
761 GACATT 766
```